

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2002, 15:28:23 ; Search time 50.82 Seconds
(without alignments)
6481.592 Million cell updates/sec

Title: US-09-910-186A-7
Perfect score: 1341
Sequence: 1 gaattcacgatggccaacaa.....ggaccgaatagtaagaattc 1341

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 383533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA: *
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq: *
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq: *
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq: *
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq: *
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq: *
6: /cgn2_6/ptodata/2/ina/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	263	19.6	1330	1 US-08-480-604A-22	Sequence 22, Appl
2	263	19.6	1330	2 US-08-405-496A-22	Sequence 22, Appl
3	263	19.6	1330	4 US-08-915-136-22	Sequence 22, Appl
4	263	19.6	1402	1 US-08-480-604A-25	Sequence 25, Appl
5	263	19.6	1402	2 US-08-405-496A-25	Sequence 25, Appl
6	263	19.6	1402	4 US-08-915-136-25	Sequence 25, Appl
7	131.4	9.8	1359	1 US-07-618-312A-3	Sequence 3, Appl
8	131.4	9.8	1359	1 US-08-280-228-3	Sequence 3, Appl
9	123	9.2	3891	1 US-08-480-604A-27	Sequence 27, Appl
10	123	9.2	3891	2 US-08-405-496A-27	Sequence 27, Appl
11	123	9.2	3891	4 US-08-915-136-27	Sequence 27, Appl
12	70.6	5.3	5163	3 US-08-700-651-1	Sequence 1, Appl
13	70.6	5.3	5163	3 US-08-928-361B-4	Sequence 4, Appl
14	70.6	5.3	5318	3 US-08-700-651-2	Sequence 2, Appl
15	70.6	5.3	5318	3 US-08-928-361B-2	Sequence 3, Appl
16	64	4.8	5511	3 US-08-928-361B-2	Sequence 2, Appl
17	64	4.8	7334	3 US-08-928-361B-1	Sequence 1, Appl
18	55.4	4.1	1359	1 US-07-618-312A-1	Sequence 1, Appl
19	55.4	4.1	1359	1 US-08-110-786A-7	Sequence 7, Appl
20	55.4	4.1	1359	1 US-08-280-228-1	Sequence 1, Appl
21	55.4	4.1	1858	1 US-08-668-381A-6	Sequence 6, Appl
22	43.4	3.2	1430	1 US-08-276-452A-25	Sequence 25, Appl
23	43.4	3.2	1430	2 US-08-798-744-25	Sequence 25, Appl
24	40.4	3.0	1690	1 US-08-276-452A-24	Sequence 24, Appl
25	40.4	3.0	1690	2 US-08-798-744-24	Sequence 24, Appl
26	38.4	2.9	198	5 PCT-US95-10668-3	Sequence 3, Appl
27	38.4	2.9	198	5 PCT-US95-10668-4	Sequence 4, Appl

28 36.8 2.7 9046 1 US-08-227-536-1 Sequence 1, Appl
29 36.8 2.7 9046 5 PCT-US95-04682-1 Sequence 1, Appl
30 36.6 2.7 6826 3 US-09-024-020B-8 Sequence 8, Appl
31 36.6 2.7 6826 4 US-09-425-043-8 Sequence 6, Appl
32 36 2.7 1149 4 US-09-422-487-6 Sequence 11, Appl
33 36 2.7 18596 4 US-09-318-448-11 Sequence 1, Appl
34 35.6 2.7 3378 1 US-07-972-791-1 Sequence 1, Appl
35 35.2 2.6 198 5 PCT-US95-10668-1 Sequence 2, Appl
36 35.2 2.6 198 5 PCT-US95-10668-2 Sequence 1, Appl
37 35 2.6 4991 4 US-09-351-200-1 Sequence 1, Appl
38 35 2.6 5977 3 US-09-024-020B-1 Sequence 1, Appl
39 35 2.6 5977 4 US-09-425-043-1 Sequence 2, Appl
40 35 2.6 6007 3 US-09-024-020B-2 Sequence 2, Appl
41 35 2.6 6007 4 US-09-425-043-2 Sequence 2, Appl
42 35 2.6 6556 3 US-09-024-020B-7 Sequence 7, Appl
43 35 2.6 6556 4 US-09-425-043-7 Sequence 7, Appl
44 35 2.6 6586 3 US-09-024-020B-43 Sequence 43, Appl
45 35 2.6 6586 4 US-09-425-043-43 Sequence 43, Appl

ALIGNMENTS

RESULT 1
US-08-480-604A-22
; Sequence 22, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESS: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027

	;	INFORMATION FOR SEQ ID NO:	22:
	;	SEQUENCE CHARACTERISTICS:	
	;	LENGTH:	1330 base pairs
	;	TYPE:	nucleic acid
	;	STRANDEDNESS:	double
	;	TOPOLOGY:	linear
	;	MOLECULE TYPE:	DNA (genomic)
	;	FEATURES:	
	;	NAME/KEY:	CDS
	;	LOCATION:	1..1314
	;	US-08-405-496A-22	

	Query Match	19.6%	Score 263;	DB 2;	Length 1330;
	Best Local Similarity	53.9%;	Pred. No. 1.7e-66;		
	Matches 639;	Conservative 0;	Mismatches 525;	Indels 21;	Gaps

Y	15	caacaatacacaattccgaatctcgtaacaaatatcatcctggaacctgcgttacaagaacaa	74
Db	24	CAC TGAATACATCAAGAACAATCATCAATACCCTCCATCTGAACTCGGTACGATCCAA	83
QY	75	caatctgatcgatctgtctgttgtaoagttgcgtctaaagttaaagtatacgaagggtgtt	129
Db	84	TCACTTGATCGACCTGCTCGGTACGCTTCCAAAATCAACATCGGTTCTAAAGTTAACTT	143
QY	130	-gaactgaatgacaagaaccaggttcaaatggacctcttcogtctaactctaaagatcogtgt	188
Db	144	CGATCCGATCGCAGCAAGAANTCAGATCCAGCTGTTCAAATCTGGAATCTTCCAAAATCGAAGT	203
QY	189	tactcgaatcagaacatcatcttcaactccgtattccctggactictgtttccctcttg	248
Db	204	TATCTTGAGATGCTATGCTATACAACTATGTAAGAAAACTTCTCACCTCTCTTCG	263
QY	249	gattcgtatcccogaaaatacaagaacgacggtatccagaattacatccccaatgaatacac	308
Db	264	GATCCGTATCCGGAAATACTTCAACTCC-----ATCTCTGAACAATGAATACAC	314
QY	309	cattccaactcatgaaagaataaactctgtgttggaagatctccatcgcgggtaacogtat	368
Db	315	CATCATCAACTGCATGGAAACAATTTCTGGTTGGAAAGTATCTCTGAACATACGGTGAAT	374
QY	369	catctggactctgatcgatatcaacggttaagaccaaatctgtattcttcgaatacaaacat	428
Db	375	CATCTGGACTCGCAGGACACTCAGGAATCAACACGCGTGTGTATTCAATACTCTCA	434
QY	429	cogtgaagacatctctgaaatacatcaatcgcgtgttcttcgttacocatcaccaataa---	485
Db	435	GATGATCAACATCTCTGACTACATCAATCGCTGGATCTTCGTTACCATCACCAACAATCG	494
QY	486	cttgaacaatgctaaaaatcatcaacggttaaacgtgaaatctaatatacgcgacatacaaga	545
Db	495	TC TGAATAACTCCAANAATCATATCAACGGCGCTCTGTATCGACCAAGAACCCGATCTCCA	554
QY	546	catccgtgaagttagctgtaacgggtgaatatcatcttcaactggacgggtgacatcgatcg	605
Db	555	TC TGGGTAACTCCAGCTTCTAANTAACATCATGTTCAACTGGAGGGTGTCTGTGACAC	614
QY	606	taccocagttcaatctggatgaaatacttctccatcttccaacacgaactgtctcagttccaa	665
Db	615	TCACCGCTACATCTGGATCAAAATACTTCAATCTGTTCGCAAAAGAACTGAACGAAAAAGA	674
QY	666	tatcgaagaacggttacaagatccagttcttaactccgaataactgaagactctggggtaa	725
Db	675	AATCAAGACCTGTACGCAACACAGTCCAATTTCTGTGATCTCTGAAAGACTTCTGGGGTGA	734
QY	726	tcocgtctgattacaacaagaatacttatatgtttcaatgcttggttaacaagaactcttaacat	785
Db	735	CTACCTGTCAGTACGCAAAACCGTACTACATGCTGAATCTGTACGATCCGCAACAATACGT	794
QY	786	caaatgaagaagactctccgggttgggtgaataactgactcgtttccaaatacaacaagaa	845
Db	795	TGACGCTCAAAATGTAGGTATCCGGGGTTTACATGTACCTGAAAGGTCGCCGCTGGTTCTGT	854

FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPD-01308
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 25:
SEQUENCE CHARACTERISTICS:
LENGTH: 1402 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1386
US-08-405-496A-25

Query Match 19.6%; Score 263; DB 2; Length 1402;
Best Local Similarity 53.9%; Pred. No. 1.7e-66;
Matches 639; Conservative 0; Mismatches 525; Indels 21; Gaps 4;
Qy 15 caacaaatacattccggaatcctggaacaataatcctgaacctggtgtacaaagacaa 74
Db 96 CACTGATATACATCAAGACATCATATACCTCCATCTGAACCTGGCTACGATCAAA 155
Qy 75 caatctgatgatgtctgtgttgacgtgtgtaagtgaagtatacagcgtgtt----- 129
Db 156 TCACCTGATCGACTGCTCGCTACGCTTCCAAATCAACATCGGTTCTAAAGTTAACTT 215
Qy 130 -gaactgaatgacaaagacagttcaactgaacctcttcgcttaactcctaagatccgtg 188
Db 216 CGATCCGATCGACAAAGATCAGATCCAGCTGTTCATCTGGAAATCTTCCAAATCGAAGT 275
Qy 189 tactcagaatcagaacatcatcttcaactcgtattcttggaactctctgtttctctgt 248
Db 276 TATCTGAAGATCTATCTGTATACACTCTATGTAGCAAACTCTCCACCCTCTCTG 335
Qy 249 gattcgtatccggaatacacaagacacggtgtatccgaataatataccacaatgaatacac 308
Db 336 GATCCGATCCGAAATACATCTCAACTCC-----ATCTCTGAACAATGAATACAC 386
Qy 309 catcataactgcatgaagaataactctgtgtggaagatctccatccgcggttaacctat 368
Db 387 CATCATCAACTGCATGGAACAAATCTGTTGGAAAGTATCTCTGAACCTACGGTGAAT 446
Qy 369 catctgactctgatcatcaacggtgaagaccacaaatctgtattctcgaatacaaat 428
Db 447 CATCTGACTCTGAGGACACTCAGGAATCAACACGCTGTTGTTATCAAAATCTCA 506
Qy 429 ccgtgaagacatctctgaatacatcaatcgtgttcttctgttaccatcccaataa--- 485
Db 507 GATGATCAACATCTCTGACTACATCAATCGTGGATCTCTGTTACCATCAACCAATCG 566
Qy 486 cctgaacaatgttaaatctacatcaacggttaactggaatcctaataccgacatcaaga 545
Db 567 TCTGAATTAATCCAAATCTACATCAACGCGCTGTGATCGACCAACCAACCGATCTCAA 626
Qy 546 catcgtgaagtctatcgctgaactgaatccttcaactggaactggtgacatcgatcg 605
Db 627 TCTGGTTAATCATCCCGTCTTATATACATCATGTTCAAACTGGACGTTGTCGTGACAC 686
Qy 606 taccagttcatctggtgaataactcttccatcttcaacacagaactgtctcagttccaa 665
Db 687 TCACCGCTACATCTGGATCAAAATCTCAATCTCTGTCGACAAAGAACTGAACGAAAAA 746
Qy 666 tatcgaagaacggttacagatccagttactcgaatacctgaactgaactcttggttaa 725
Db 747 AATCAAGAGCTGTGACGACCAACAGTCCAATCTGTTGATCTCTGAAGACTTCTGGGGTGA 806

Qy 726 tccgctgatgtacacaaagaataactatgttcaatgctggtgaacgaactcttaccat 785
Db 807 CTACCTGCAGTACGACAAACCGTACTACTGAATCTGTACGATCCGAACAATACTGT 866
Qy 786 caaactgaagaagactctccggttggtagaaactcctgactcgttccaaatacacaaccgaa 845
Db 867 TGAGTCAACAATGTAGGTATCCGCGGTATACATGTACCTGAAAGGTCGCGGTGTTCTGT 926
Qy 846 ctctaaatacatcaactacacgacactgacatcggtgaaagttcatcctcgcgcaaa 905
Db 927 TATGACTACCAACATCTACCTGAACTCTTCCTCTACCGTGGTACCAATTCATCATCAA 986
Qy 906 atctaactctcagtcctcaatgatgatcgtacgtacgtaaaagaagactacatctcctgga 965
Db 987 GAAATACGCGTCTGTAAACAGGACAAATATCGTTCCGCAACAATGATCGTGATACATCAA 1046
Qy 966 ctcttcaacctgaatcaggaatggcgtgtatatacactacacgaagtacttcaagaagaaga 1025
Db 1047 TGTGTAGTTAAGAACAAGAATACCGTCTGGCTTACCAATGCTTCTCAGGCTGGGTAGA 1106
Qy 1026 agaaagcttttctcgtccgactctgattccgacgaaactctacaacaccatccagat 1085
Db 1107 AAGATCTTGTCTCTCTGAAATCCCGGACGTTGGTAATCTGTCTCAGGTAGTTGTAAT 1166
Qy 1086 caaagaatacagacagacagccgaactactcttgcagctgctgttcaagaagaagatgaaga 1145
Db 1167 GAAATCCAAAGAACGACGACGGGTA---TCACTAACAAATGCAAAATGAATCTGCAGGACAA 1223
Qy 1146 atctaactgaagaatcgtgtgatcgtgtatccacacggtttctacga 1190
Db 1224 CAATGGTAACGATATCGGTTTCATCGGTTTCCACCAGTTCAACAA 1268

RESULT 6

US-08-915-136-25
; Sequence 25, Application US/08915136
; Patent No. 6290960
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: FADHIE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,136
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/480,604
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907

NAME: Crawford Mr, Arthur R
REGISTRATION NUMBER: 25,327
REFERENCE/DOCKET NUMBER: 510-51
TELECOMMUNICATION INFORMATION:
TELEPHONE: 0101 703 8750400
TELEFAX: 0101 703 5253468
TELEX: 200797 NIXN UR
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 1359 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Clostridium tetani
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1356
US-07-618-312A-3

Query Match 9.8%; Score 131.4; DB 1; Length 1359;
Best Local Similarity 47.6%; Pred. No. 2e-28;
Matches 633; Conservative 0; Mismatches 636; Indels 60; Gaps 6;
QY 40 aacaatcatctcgaacgtggtgtaacaaagacaacaatctgcatctgtctggttac 99
DB 58 AAGTCTACCATTTGAATCTGGACATCAACACGATATTATCTCCGACATCTCTGGTTTC 117
QY 100 ggtgctaaagtgaagatatacagcgtgttgtaactgaatgacaagaacacgttcaactg 159
DB 118 AACTCTCTGTTATCATCATCAGATGCTCAATTTGGTCCGGGCAATCAACGGCAAGCT 177
QY 160 acctcttcgcgttaac-----tctaagatcgcgtgttactcagaatcagaacatcatc 210
DB 178 ATCCACTGGTTPAACACAGNATCTTCTGAAGTTATCGTGCAACAGGCCATGGACATCGAA 237
QY 211 ttaacacgttatctcgtgactctctgttctctctgtgattcgatcccgaaatacaag 270
DB 238 TACAAGGACATGTTTCAACAACATTCACCGTTAGCTTCTGGCTGGCGCTTCGGAAGATTCT 297
QY 271 aacgacggtatccagatattacccaacatgaatacaccatcaactcaactgcatgaagaat 330
DB 298 GCTTCCACCTGGACATAGTACGGCACTACGAGTATCTCCATCATCATAGCTCTATGAAGAA 357
QY 331 aa-----ctctggttggaagatctccatccgcggttaacccgtatcatctg 375
DB 358 CACTCCCTGTCCATCGCTCTGTTGGTCTGTTTCCCTGAAGGGTAACAACCTGATCTGG 417
QY 376 actctgatgatataacggtgaagacaacaatctgtattcttctgaatacaacatccgtgaa 435
DB 418 ACTCTGAAAGACTCCCGGGCGGAAGTTCGTAGATCACTTTCCGCGACCTCGCGGACAAG 477
QY 436 gacatctctgaatacatcatcgtggtctctctgttaccatcac---caataacctgaaac 492
DB 478 TTCACGGGTACCTGGCTACAAATAGGGTTTTCATCACTATCACTAACGATCTGTCGTCT 537
QY 493 aatgctaaatctacatcaacggttaactggaatctaataccgacatcaaaagacatccgt 552
DB 538 TCTGCTAACTGTACATCAACGGCGTTCTGATGGGCTCCGCTGAATCACTACTGTCTGGGC 597
QY 553 gaagtattcgtactgaaggtgaatcatcttcaaaactggacggtgacatcgatcgatcccgag 612
DB 598 GCTATCCGTGAGGACAAACATCACTCTTAAGCTGGACCGTTGCAACCAACAACACACAG 657
QY 613 ttcatctggtgaataactcttccatctcttcaacacccggaactctcagtcaccaatcgaa 672
DB 658 TAGGTATCATCGACAAGTTCGTATCTTCTGCAAGACTGAACCCGGAAGAGATCGAA 717
QY 673 gaacggtacaagatccagttcttactccgaataacctgaaagacttctggtggttaacccgtg 732

DB 718 AAACGTGTATACCAGTACCTGTCTATCACCTTCTCGGTGACTTCTGGGGTAACCCGCTG 777
QY 733 atgtacaacaagaataactatattgttcaatgctggtgaacaagaactcttacatcaactg 792
DB 778 CGTTACGACACCGAATATTACCTGATCCCGGTAGCTTCTAGCTCTTAAGACGCTACAGTCCG 837
QY 793 aagaagaactctccggttggtgaaatccctgactcgttccaataacacaccgaactctctaaa 852
DB 838 AAAACATCACTGACTACATGTACCTGACCAACGCGCGCTCTACACTACCGTAAACTG 897
QY 853 tacatcaactaccgcgacgtgtacatcggtgaaagttaactatccgtccgtaacttaac 912
DB 898 AACATCTACTACCGAGCTCTGTACAACGGCTGAAATTCATCAACACGCTACACTCCG 957
QY 913 tctcagtcocatcaatgatgacatcgtacgttaagaagaactacatctaccctggactcttc 972
DB 958 AACACGAAATCGATTTCTGTTAAATCTGGTGACTTCATCAAACTGTGCGTTCTTTAC 1017
QY 973 aacctgaatcagggaatggcggtgtatcacacctac-----aagtacttcaagaagaagaa 1026
DB 1018 AACACACGACACATCGTTGGTTACCGGAAGACGGTAACGCTTTCACACACCTGGAC 1077
QY 1027 gaaaagctttctcgtcgcgtatctctgattccgaggaactctacaacacatccagatc 1086
DB 1078 AGAATTTCTGCGTGTGTTACAACGCTCCGGGTATCCCGCTGTACAAAAAATGGAAGCT 1137
QY 1087 aaagaatacagacaacagcgcgacactactcttgcacgctgctgttcaagaagaatgaagaa 1146
DB 1138 GTTAAACTCGTGACCTGAAACCTACTCTGTGTTCACTGAGTGAACCTGTACGACGACAAAAC 1197
QY 1147 tctactgacgaaatcgggtcgtgattccaccggtttctacgaatctggtatctgtattc 1206
DB 1198 GCTTCTC-----TGGGTCTGGTTGGTACCCACCAACGCTCAGATCGGTAAAC 1242
QY 1207 gaagaatacaaaagactactctctcatctccaaatggttacctgaaggaagttaacgcaaa 1266
DB 1243 GACCCGAACCGTCACATCTCTGATCGCTTCTTAACCTGGTACTTCAACACCTGAAAGACAAA 1302
QY 1267 cgtacaactcgaactgaaggttgcaattggcagttcattccgaaagaacgaaggttgacc 1326
DB 1303 -----ATCTGGGTTGCGACTGTGTTCTCCGACCGATGAAGGTTGGACC 1350
QY 1327 gaatagtaa 1335
DB 1351 AACGACTAA 1359

RESULT 8
US-08-280-228-3
; Sequence 3, Application US/08280228
; Patent No. 5571694
; GENERAL INFORMATION:
; APPLICANT: Makoff Dr, Andrew J
; APPLICANT: Romanos Dr, Michael A
; APPLICANT: Clare Dr, Jeffrey J
; APPLICANT: Fairweather Dr, Neil F
; TITLE OF INVENTION: VACCINES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 No. 5571694th Glebe Road
; CITY: Arlington,
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/280,228


```
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPND-01308
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3891 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3888
; US-08-405-496A-27

Query Match          9.2%; Score 123; DB 2; Length 3891;
Best Local Similarity 50.2%; Pred. No. 9.5e-26;
Matches 396; Conservative 0; Mismatches 375; Indels 18; Gaps 3;

Qy 20 aatacaattccgaaatcgcgaacataatcctcgaactcgttcaacagcgttcaacagacaacatc 79
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2603 AATATATTAAAGATATTATTATTAATCTCTATATTGAATTTAAGATATGAAGTAAATCAT 2662

Qy 80 tgaatgatctctcgtggttacggtgtaaa-----gttgaagtatacagcgggtgttgaaac 133
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2663 TAATAGACTTATCTAGGTATGATCAAAATAAATAATATTGGTAGTAAGTAAATTTTGATC 2722

Qy 134 tgaatgacaagaacagcttcaactccttcgccttaactcctaagatccggtgttaactc 193
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2723 CAATAGATAAAATCAAAATTCATTTTAAATTTTGAAGAGTAGTAAATTCAGGTAAATTT 2782

Qy 194 agaatacagaacatcatcttcaactcgcgtattccctggaactctcttccctctggattc 253
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2783 TAAAAAATGCTATTGTATATATAGTATGTATGAATAATTTAGTACTAGCTTTTGGATAA 2842

Qy 254 gtatcccgaaatacagaacgacgggtatccagaattatcaccaatgaatacacatca 313
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2843 GAATTCCTAAGTATTTTAAACAGTATAAGTCTAAAT-----AATGAATATACAATAA 2893

Qy 314 tcaactgcatgaagaataaactctggttggaagatctccatccgcgcgggtaacgcgtatcatct 373
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2894 TAAATGTATGGAATAAATTCAGGATGGAAGTAGTACTTAATTTATGTTGGTGAATAATCT 2953

Qy 374 ggactctgatgatcaacggtgaagacaaactctgtattcttgaatacacaacatccggtg 433
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2954 GGACTTTACAGGATACTCAGGAATAAACAAGAGTAGTTTTTAAATACAGTCAAAATGA 3013

Qy 434 aagacatctctgaatacactaatcgtgttcttccatccatccaccataaa---cctga 490
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3014 TTAATATATACAGATTATATAACAGATGGATTTTGTAACTATCACTAATAATAGATTAA 3073

Qy 491 acaatgctctcaactacacggttaactggaatcttaataccgacatcaaaagacatcc 550
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3074 ATAACTTAAATTTATATAATGGAAGATTAAATAGATCAAAAACCAATTTCAAAATTTAG 3133

Qy 551 gtgaagttatcgtcaacggtgaaatcattcctcaactggacggtgacatcgatcgatccc 610
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3134 GTAATATTCATGCTAGTAAATATAATATATATGTTTAAATAGATGGTGTAGAGATACACATA 3193

Qy 611 agttcatctggatgaataactctctccatcttcaacaccggaactgctcagtcaccaatcgc 670
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 3194 GATATATTGGATAAAATATTTTAATCTTTTGTATAAGGAATTAATAAGAAAAAATCA 3253

Qy 671 aagaacggtaceagatccagctcttactccgaataactcgaagactctcggggttaaccgcg 730
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3254 AAGATTATATGATGATAATCAATCAAAATTCAGGTATTTTAAAGACTTTTGGGGTGATTATT 3313

Qy 731 tgaatgacaacaaagaataactatattgttcaatgctcgtgtaacaagaactcttaccatcaaac 790
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3314 TACAATATGATAAACCATACTATATGTTAAATTTATATGATCCAAATAATATGTCGATG 3373

Qy 791 tgaagaagaag 799
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 3374 TAAATAATG 3382

RESULT 11
US-08-915-136-27
; Sequence 27, Application US/08915136
; Patent No. 6290960
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,136
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/480,604
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPND-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3891 base pairs
; TYPE: nucleic acid
```


OM of: US-09-910-186A-8 to: Issued_Patents_NA:* out_format : pfs
Date: Sep 2, 2002 4:01 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL=framet_p2n.model -DEV=xlh
-Q=/cgn2_1/USPRO_spool/US09910186/runat_29082002_134725_16013/app_query.fasta_1.505
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blsum62
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09910186.@CGN1_1_51 -NCPU=6 -ICPU=3 -LONGLOG.
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:
Query: US-09-910-186A-8
Query length: 440
Database: Issued_Patents_NA:*
Database sequences: 383533
Database length: 122816752
Search time (sec): 47.780000

Sequence	Strd Orig	ZScore	EScore	Len	Documentation
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-480-604A-22		831.50	1619.44	1.6e-82	1330
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-405-496A-22		831.50	1619.44	1.6e-82	1330
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-915-136-22		831.50	1619.44	1.6e-82	1330
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-480-604A-25		831.50	1618.83	1.8e-82	1402
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-405-496A-25		831.50	1618.83	1.8e-82	1402
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-915-136-25		831.50	1618.83	1.8e-82	1402
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-480-604A-27		831.50	1607.12	7.9e-82	3891
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-405-496A-27		831.50	1607.12	7.9e-82	3891
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-915-136-27		831.50	1607.12	7.9e-82	3891
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-07-618-312A-1		595.00	1150.30	2.2e-56	1359
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-280-328-1		595.00	1150.30	2.2e-56	1359
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-07-618-312A-3		594.00	1148.31	2.8e-56	1359
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-280-328-3		594.00	1148.31	2.8e-56	1359
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-668-381A-6		594.00	1144.73	4.5e-56	1858
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-110-786A-7		575.00	1110.64	3.6e-54	1359
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-446-855A-1		163.00	272.21	1.8e-07	8920
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-150-741-1		163.00	272.21	1.8e-07	8920
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-09-316-083-2		137.50	242.65	7.9e-06	1431
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-276-452A-24		128.00	221.90	0.0001	1690
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-798-744-24		128.00	221.90	0.0001	1690
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-568-459A-3		126.00	206.69	0.0008	4507
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-487-826B-3		126.00	206.69	0.0008	4507
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-686-968C-1		124.50	207.77	0.0007	3164
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-480-604A-9		123.50	196.52	0.0029	7101
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-405-496A-9		123.50	196.52	0.0029	7101
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-915-136-9		123.50	196.52	0.0029	7101
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-323-170B-1		123.50	193.01	0.0046	9636
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-954-441-1		123.50	193.01	0.0046	9636
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-853-659A-8		123.50	187.55	0.0093	15512
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-853-659A-8		123.50	187.55	0.0093	15512
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-853-659A-63		123.50	187.55	0.0093	15512
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-853-659A-66		123.50	187.55	0.0093	15512
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-853-659A-2		123.50	182.22	0.0184	24701
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-853-659A-3		123.50	182.22	0.0184	24701
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-853-659A-60		123.50	182.22	0.0184	24701
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-853-659A-61		123.50	182.22	0.0184	24701
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-413-118-117		120.50	206.57	0.0008	1760
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-473-446-117		120.50	206.57	0.0008	1760
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-480-604A-5		120.00	188.02	0.0087	8133
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-405-496A-5		120.00	188.02	0.0087	8133

/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-915-136-5 + 120.00 188.02 0.0087 8133
/cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-951-871-1 + 117.00 194.56 0.0038 2739
/cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-851-843A-1 + 117.00 192.49 0.0049 3279
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-974-349A-109 + 117.00 192.49 0.0049 327
/cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-854-050-1 + 117.00 192.49 0.0049 3279
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-480-604A-22
seq_documentation_block:
; Sequence 22, Application US/08480604A
; Patent No. 5736139
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; TITLE OF INVENTION: PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/480,604A
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/422,711
; FILING DATE: 14-APR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 22:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1330 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1314
US-08-480-604A-22


```

1258 CAGTTCACAAATATCGCT.....AAACTGGT 128
407  sileSerLysTrpTyrLeuLysGluValLysArgLysProTyrAsnLeuL 424
      |||||:||||| :|||:|||||:|||||:
1284  TCGTTCCAACTGGTACAACTCGTCAGATCGAAGCTTCC.....TCTCGCA 1327
424  ysLeuGlyCysAsnTrpGlnPheIleProLysAspGluGlyTrpThrGlu 440
      |||||:|||||:|||||:|||||:|||||:|||||
1328  CTCCTGGTGTCTCTTGGGAGTTTCATCCCGTTGATCAGCGTTGGGTGAA 1377

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-405-496A-25

seq_documentation_block:
; Sequence 25, Application US/08405496A
; Patent No. 5919665
; GENERAL INFORMATION:
; APPLICANT: WILLIAMS, JAMES A.
; TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
; TITLE OF INVENTION: NEUROTOXIN
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: USA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/405.496A
; FILING DATE: 16-MAR-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01308
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1402 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1386
; US-08-405-496A-25

alignment_scores:
  Quality: 831.50      Length: 450
  Ratio: 2.607        Gaps: 10
  Percent Similarity: 70.889      Percent Identity: 40.000

```

alignment_scores:		
Quality:	831.50	Length: 450
Ratio:	2.607	Gaps: 10
Percent Similarity:	70.889	Percent Identity: 40.000


```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPHD-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 25:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1402 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1386
; US-08-915-136-25

```

```

alignment_scores:
  Quality: 831.50      Length: 450
  Ratio: 2.607         Gaps: 10
Percent Similarity: 70.889 Percent Identity: 40.000

```

```

alignment_block:
US-09-910-186A-8 x US-08-915-136-25

```

```

Align seg 1/1 to: US-08-915-136-25 from: 1 to: 1402

```

```

4 LysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeuArgTyrLy 20
100 GAATACATCAAGAACATCATCAATACCTCCATCCTGGAACCTGCGCTACGA 149
20 sAspAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluValT 37
150 ATCCAATCACCTGATCGACCTGCTCGCTACGCTCCAAAATCAACATCG 199
37 yrAspGlyValGluLeuAsn.....AspLysAsnGlnPheLysLeuThr 51
200 GTTCTAAAGTTAACTTCGATCCGATCGACAAAGAATCAGATCCAGCTGTC 249
52 SerSerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePh 68
250 AATCTGGAACTCTCCAAATCGAAGTTATCTCGAAGAATGCTATCGTATA 299
68 eAsnSerValPheLeuAspPheSerValSerPheTrpIleArgIleProL 85
300 CAACTCTATGTACGAAAACTTCTCCACCTCTCTGGATCGGTATCCCGA 349
85 ysTyrLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIle 101
350 AATACTTCAAC.....TCCATCTCTGGAACAATGAATCAACCATC 390
102 IleAsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgGlyAs 118
391 ATCAACTGCATGGAAACAATTCGTGGTGGAAAGTATCTCTGAACATCGG 440

```

```

118 nArgIleIleTrpThrLeuIleAspIleAsnGlyLysThrLysSerValP 135
441 TGAATCATCTGGACTTCGAGGACACTCAGGAAATCAACAGCGTGTG 490
135 hePheGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTrp 151
491 TATTCAAAATACTCTCAGATGATCAACATCTCTGACTACATCAATCGCTG 540
152 PhePheValThrIleThrAsnAsn....LeuAsnAsnAlaLysIleTyrIl 167
541 ATCTTCGTTACCATCACCACAACTCGTCTGAATACTCCAAAATCTACAT 590
167 eAsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGluValI 184
591 CAACGGCGCTGATCGACACAGAACCGATCTCCAATCTGGGTAACATCC 640
184 leAlaAsnGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThr 200
641 ACGCTTCTAATAACATCATCTTCAAACTGGACGGTTGTCGTGACACTCAC 690
201 GlnPheIleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGl 217
691 CGCTACATCTGGATAAATACTTCAATCTGTTCGACAAAGAACTGAACGA 740
217 nSerAsnIleGluArgTyrLysIleGlnSerTyrSerGluTyrLeuL 234
741 AAAGAAATCAAGACCTGTACGACAAACCACTCCAATTTCTGTATCCTGA 790
234 ysAspPheTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPhe 250
791 AAGACTTCTGGGTGACTACCTGCTGACGACGACAAACCGTACTACATGCTG 840
251 AsnAlaGlyAsnLysAsnSerTyrIleLysLeuLys..... 263
841 AATCTGTACGATCCGACAAATACGTTGACGTCAACAATGTAGGTATCCG 890
264 .....AspSerProValGlyIleLeuLeuThrArgSerL 275
891 CGGTTACATGTACCTGAAAGCTCCGCGTGTCTGTTATGACTACCAACA 940
275 ysTyrAsnGlnAsnSerLysTyrIleAsnTyrArgAspLeuTyrIleGly 291
941 TCTACCTGAACCTCTCC.....CTGTACCGTGTGT 969
292 GluLysPheIleIleArgArgLysSerAsnSerGlnSerIleAsnAspAs 308
970 ACCAAATCATCATCAGAAATACGCTCTGCTGTAAC.....AAGGACAA 1013
308 pIleValArgLysGluAspTyrIleTyrLeuAspPhePheAsnLeuAsnG 325
1014 TATCGTTCCGCAACAATGATCGTGATATACATCAATGTTGTAGTTAAGA 1063
325 lnGluTrpArgValTyrThrTyrLysTyrPheLysLysGluGluLys 341
1064 AAGAAATACCGTCTGGCTACCAATGCTTCTCAGGCTGGGTGAGAAAAGATC 1113
342 LeuPheLeuAlaProIleSerAspSerAspGluLeuTyrAsnThrIleGl 358
1114 TTGCTGCTCTGGAATCCCGGACGTTGGTAATCTGCTCAGGTAGTTGT 1163
358 nIleLysGluTyrAspGluGlnPro...ThrTyrSerCysGlnLeuLeup 374
1164 AATGAAATCCAAAGACGACGAGGTATCAGTATCGGTTTCATCGGTTCCAC 1209
374 heLysLysAspGluSerThrAspGluIleGlyLeuIleGlyIleHis 390
1210 ..AATCTCAGGACAACAATGTTACGATATCGGTTTCATCGGTTCCAC 1257
391 ArgPheTyrGluSerGlyIleValPheGluGluTyrLysAspTyrPheCy 407
1258 CAGTTCAACAATATCGCT.....AACTGGT 1283
407 sIleSerLysTrpTyrLeuLysGluValLysArgLysProTyrAsnLeuL 424

```

1284 TGCTTCCAAGTGTACAACTGCTCAGTGAACGTTCC.....TCTCGCA 1327
||||:||||| :||||:||||| :|||

424 ysluGlyCysAsnTrpGluPheIleProIysaspGluGlyTrpThrGlu 440
||||:||||| :||||:||||| :||||:||||| :|||

1328 CTCGGGTGCTCTGGAGATTATCCCGGTTGATGACGGTGGGGTAA 1377

seq_name: /cqn2_6/ptodata/2/ina/5A_COMB.seq:us-08-480-604A-27

seq_documentation_block:

; Sequence 27, Application US/08480604A

; Patent No. 5736139

; GENERAL INFORMATION:

; APPLICANT: KINK, JOHN A.

; APPLICANT: THALLEY, BRUCE S.

; APPLICANT: PADHYE, NISHA V.

; APPLICANT: FIRCA, JOSEPH R.

; APPLICANT: STAFFORD, DOUGLAS C.

; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND

; PREVENTION OF C. DIFFICILE DISEASE

; NUMBER OF SEQUENCES: 32

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: MEDLEN & CARROLL, LLP

; STREET: 220 MONTGOMERY STREET, SUITE 2200

; CITY: SAN FRANCISCO

; STATE: CALIFORNIA

; COUNTRY: UNITED STATES OF AMERICA

; ZIP: 94104

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/480,604A

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 424

; APPLICATION DATA:

; APPLICATION NUMBER: US 08/422,711

; FILING DATE: 14-APR-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/405,496

; FILING DATE: 16-MAR-1995

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/329,154

; FILING DATE: 25-OCT-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/161,907

; FILING DATE: 02-DEC-1993

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/985,321

; FILING DATE: 04-DEC-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 07/429,791

; FILING DATE: 31-OCT-1989

; ATTORNEY/AGENT INFORMATION:

; NAME: INGOLIA, DIANE E.

; REGISTRATION NUMBER: 40,027

; REFERENCE/DOCKET NUMBER: OPHD-01763

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (415) 705-8410

; TELEFAX: (415) 397-8338

; INFORMATION FOR SEQ ID NO: 27:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 3891 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: linear

; MOLECULE TYPE: DNA (genomic)

; FEATURE:

; NAME/KEY: CDS

; LOCATION: 1..3888

US-08-480-604A-27

alignment_scores:

Quality: 831.50 Length: 450

Ratio: 2.607 Gaps: 10

Percent Similarity: 70.889 Percent Identity: 40.000

alignment_block:

US-09-910-186A-8 x US-08-480-604A-27 ..

Align seg 1/1 to: US-08-480-604A-27 from: 1 to: 3891

4 LysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeuArgTyrLY 20
:||||:||||| :||||:||||| :||||:||||| :|||

2602 GAATATATTAAAGATATTATTACTTCTATATTGAATTAAGATATGA 2651
:||||:||||| :||||:||||| :||||:||||| :|||

20 saspAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluValT 37
:||||:||||| :||||:||||| :||||:||||| :|||

2652 AAGTAATCAATTAATAGACTTATCTAGGTATGCATCAAAATAAATAT 2701
:||||:||||| :||||:||||| :||||:||||| :|||

37 yrAspGlyValGluLeuAsn.....AspLysAsnGlnPheLysLeuThr 51
:||||:||||| :||||:||||| :||||:||||| :|||

2702 GTAGTAAAGTAAATTTTGTCCCAATAGATAAAATCAAAATCAATTT 2751
:||||:||||| :||||:||||| :||||:||||| :|||

52 SerSerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePh 68
:||||:||||| :||||:||||| :||||:||||| :|||

2752 AATTTAGAAAGTAGTAAATTTAGGTAAATTTAAAAAATGCTATTGTAT 2801
:||||:||||| :||||:||||| :||||:||||| :|||

68 eAsnSerValPheLeuAspPheSerValSerPheTrpIleArgIleProL 85
:||||:||||| :||||:||||| :||||:||||| :|||

2802 TAATAGTATGTATGAAATTTTACTAGTCTTGGATGAAGAATTCCTA 2851
:||||:||||| :||||:||||| :||||:||||| :|||

85 ysrLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIle 101
:||||:||||| :||||:||||| :||||:||||| :|||

2852 AGTATTTTAACT.....AGTATAAGTCTAAATAATGAATATACAA 2892
:||||:||||| :||||:||||| :||||:||||| :|||

102 IleAsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgGlyAs 118
:||||:||||| :||||:||||| :||||:||||| :|||

2893 ATAAATGTATGGAAATAATTTAGGATGGAAGATATCACTTAATATTG 2942
:||||:||||| :||||:||||| :||||:||||| :|||

118 narGileIleTrpThrLeuIleAspIleAsnGlyLysThrLysSerValP 135
:||||:||||| :||||:||||| :||||:||||| :|||

2943 TGAATAAATCTGGACTTTACAGGATACCTCAGAAATAAAACAAGAGTAG 2992
:||||:||||| :||||:||||| :||||:||||| :|||

135 hePheGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTrp 151
:||||:||||| :||||:||||| :||||:||||| :|||

2993 TTTTAAATACAGTCAAAATGATTAATATATATATATATAAACACAGAT 3042
:||||:||||| :||||:||||| :||||:||||| :|||

152 PhePheValThrIleThrAsnAsn.....LeuAsnAsnAlaLysIleTyr 167
:||||:||||| :||||:||||| :||||:||||| :|||

3043 ATTTTGTAACTATCCTAATAATAGATTAATTAACCTAAATTAATATAT 3092
:||||:||||| :||||:||||| :||||:||||| :|||

167 eAsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGluValI 184
:||||:||||| :||||:||||| :||||:||||| :|||

3093 AAATGGAAGATTAATAGATCAAAACCAATTTCAAAATTTAGGTAATATC 3142
:||||:||||| :||||:||||| :||||:||||| :|||

184 leaIleAsnGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThr 200
:||||:||||| :||||:||||| :||||:||||| :|||

3143 ATGCTAGTAATAATAATGTTTAAATTTAGATGTTGTAGAGATACACAT 3192
:||||:||||| :||||:||||| :||||:||||| :|||

201 GlnPheIleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerG 217
:||||:||||| :||||:||||| :||||:||||| :|||

3193 AGATATATTTGGATAAATAATTTTAACTCTTTTGTATGAAGGAATTAATGA 3242
:||||:||||| :||||:||||| :||||:||||| :|||

217 nSerAsnIleGluArgTyrLysIleGlnSerTyrSerGluTyrLeuL 234
:||||:||||| :||||:||||| :||||:||||| :|||

3243 AAAAGAAATCAAGATTTTATATGATATCAATCAAAATTCAGGTATTTAA 3292
:||||:||||| :||||:||||| :||||:||||| :|||

234 ysaspPheTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPhe 250
:||||:||||| :||||:||||| :||||:||||| :|||

3293 AAGATTTTGGGTGATTATTACAAATATGATAAACCATATATATATGTTA 3342
:||||:||||| :||||:||||| :||||:||||| :|||

251 AsnAlaGlyAsnLysAsnSerTyrIleLysLeuLysLys..... 263

```

|||
3343 AATTATATGATCCAAATAAATATCGTAAATAATGTAGGTATTAG 3392
264 .....AspSerProValGlyGluLeuLeuThrArgSerL 275
3393 AGGTTATATGATCTTAAGGGCTAGAGGTAGCGTAATGACTACAAACA 3442
275 ySTyrAsnGlnAsnSerLysTyrIleAsnTyrArgAspLeuTyrIleGly 291
3443 TTTATTAAATTCAGT.....TTGTATAGGGGG 3471
292 GluLysPheIleIleArgLysSerAsnSerGlnSerIleAsnAspAs 308
3472 ACAAAATTTATTAAAAAATATGCTTCGCAAT.....AAGATAA 3515
308 pIleValArgLysGluAspTyrIleTyrLeuAspPheAsnLeuAsnG 325
3516 TATTGTTAGAAATAATGATCGTGTATATATTAGTAGTTAAATA 3565
325 InGluTyrArgValTyrThrTyrLysTyrPheLysLysGluGluLys 341
3566 AAGATATAGGTTACTACTAATGTCATCACAGGCGGTAGAAATA 3615
342 LeuPheLeuAlaProIleSerAspGluLeuTyrAsnThrIleG1 358
3616 CTAAGTGCAATTAGAAATACCTGATAGGAAATCTAAGTCAAGTAGTAGT 3665
358 nIleLysGluTyrAspGluGlnPro...ThrTyrSerCysGlnLeuLeuP 374
3666 AATGAAGTCAAAAATGATCAAGGAATAACAAATAAATGCAAAATG... 3711
374 heLysLysAspGluSerThrAspGluIleGlyLeuIleGlyIleHis 390
3712 ..AATTCAAGATAATATGGGAATGATAGGCTTTATAGGATTTCAT 3759
391 ArgPheTyrGluSerGlyIleValPheGluGluTyrLysAspTyrPheCy 407
3760 CAGTTTAATAATAGCT.....AACTAGT 3785
407 sIleSerLysTyrTyrLeuLysGluValLysArgLysProTyrAsnLeuL 424
3786 AGCAAGTAATGGTATAATAGCAAAATAGAAATCT.....AGTAGGA 3829
424 ySLeuGlyCysAsnTyrGlnPheIleProLysAspGluGlyTyrThrGlu 440
3830 CTTTGGGTGCTCAGGAATTTATCTCTGATAGATGATGGATGGGGAGNA 3879
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-405-496A-27
seq_documentation_block:
Sequence 27, Application US/08405496A
Patent No. 5919665
GENERAL INFORMATION:
APPLICANT: WILLIAMS, JAMES A.
TITLE OF INVENTION: VACCINE FOR CLOSTRIDIUM BOTULINUM
COMPUTER: IBM PC compatible
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESSEE: MEDLEN & CARROLL, LLP
STREET: 220 MONTGOMERY STREET, SUITE 2200
CITY: SAN FRANCISCO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/405,496A
FILING DATE: 16-MAR-1995
CLASSIFICATION: 424
```

```

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/329,154
FILING DATE: 25-OCT-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/161,907
FILING DATE: 02-DEC-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/985,321
FILING DATE: 04-DEC-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/429,791
FILING DATE: 31-OCT-1989
ATTORNEY/AGENT INFORMATION:
NAME: INGOLIA, DIANE E.
REGISTRATION NUMBER: 40,027
REFERENCE/DOCKET NUMBER: OPHD-01308
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 705-8410
TELEFAX: (415) 397-8338
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 3891 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..3888
US-08-405-496A-27

alignment_scores:
Quality: 831.50 Length: 450
Ratio: 2.607 Gaps: 10
Percent Similarity: 70.889 Percent Identity: 40.000

alignment_block:
US-09-910-186A-8 x US-08-405-496A-27 ..
Align seg 1/1 to: US-08-405-496A-27 from: 1 to: 3891
4 LysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeuArgTyrL 20
2602 GAATATATTAAGATATATTAACTACTTCTATATGAATTAAGATATGA 2651
20 sAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluValT 37
2652 AAGTAATCATTTAATAGACTTATCTAGGTATGTCATCAAAAATAATTTG 2701
37 yRAspGlyValGluLeuAsn.....AspLysAsnGlnPheLysLeuThr 51
2702 GTAGTAAGTAATAATTTTGTATCAATAGATAAAATCAAAATCAATATT 2751
52 SerSerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIlePhe 68
2752 RATTAGAAAGTAGTAAATTCAGGCTAATTTTAAAAAATGCTATTGTATA 2801
68 eAsnSerValPheLeuAspPheSerValSerPheTrpIleArgIleProL 85
2802 TAAATAGTATGTAGAAAATTTTGTACTAGCTTTTGGATAAGAAATTCCTA 2851
85 ySTyrLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIle 101
2852 AGTATTTTAAC.....AGTATAAGTCTAAATAATGAATATACATA 2892
102 IleAsnCysMetLysAsnAsnSerGlyTyrPlyIleSerIleArgGlyAs 118
2893 ATAATTTGTATGGAATAATTCAGGATGGAAGATCATCATTAAATATGG 2942
118 nArgIleIleTrpThrLeuIleAspIleAsnGlyLysThrLysSerValP 135
2943 TGAATAATCTGGACTTTTACAGGATACTCAGGAAATAAAACAAGAGTAG 2992
```

```
135 hePheGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTrp 151
|||||
2993 TTTTAAATACAGTCAAAATGATTAAATATATACAGATTATATAACAGATGG 3042
|||||
152 PhePheValThrIleThrAsnAsn...LeuAsnAsnAlaLysIleTyrIle 167
|||||
3043 ATTTTGTAACTACTACTATAATAGATTAAATAAATACTTAAATTTATAT 3092
|||||
167 eAsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGluVali 184
|||||
3093 AAATGAGAGTAATAGATCAAAACCAATTTCAATTTAGGTAATATTC 3142
|||||
184 leaIaAsnGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThr 200
|||||
3143 ATGCTAGTAATAATAATGTTTAAATTTAGATGGTTGTAGAGATACACAT 3192
|||||
201 GlnPheIleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGI 217
|||||
3193 AGATATATTGGGATAAAATATTTAATCTTTTGTAAAGGAATTAATGA 3242
|||||
217 nSerAsnIleGluIuArgTyrLysIleGlnSerTyrSerGluTyrLeuL 234
|||||
3243 AAAGAAATCAAGATTTATGATGATATCAATCAATTCAGGTATTTTAA 3292
|||||
234 ysAspPheTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPhe 250
|||||
3293 AAGACTTTTGGGGTGATTTTACAATATGATAAACCACATATATATGTTA 3342
|||||
251 AsnAlaGlyAsnLysAsnSerTyrIleLysLeuLysLys..... 263
|||||
3343 AATTTATATGATCCAAATAATATGTCGATGTAATAATGTAGTATTAG 3392
|||||
264 .....AspSerProValGlyGluIleLeuThrArgSerL 275
|||||
3393 AGTTTATATGATCTTAAAGGGCTAGAGGTAGCGTAATGACTACAAACA 3442
|||||
275 ysTyrAsnGlnAsnSerLysTyrIleAsnTyrArgAspLeuTyrIleGly 291
|||||
3443 TTTTATTAAATTCACCT.....TTGTATAGGGG 3471
|||||
292 GluLysPheIleIleArgLysSerAsnSerGlnSerIleAsnAspAs 308
|||||
3472 ACAAAATTTTATAAAAAAATATGCTTCTGGAAT.....AAAGATAA 3515
|||||
308 pIleValArgLysGluAspTyrIleTyrLeuAspPheAsnLeuAsnG 325
|||||
3516 TATGTTAGAAATATGATCGGTATATATATATGATAGTAGTTAAATAA 3565
|||||
325 lngluTrpArgValTyrThrTyrLysTyrPheLysLysGluGluLys 341
|||||
3566 AAGAATATAGTTAGTACTACTATGTCATCACAGCGCAGCGTAGAAAAA 3615
|||||
342 LeuPheLeuAlaProIleSerAspSerAspGluLeuTyrAsnThrIle 358
|||||
3616 CTAAGTCATTAGAAATACCTCTAGGAAATCTAAGTCAAGTAGTAGT 3665
|||||
358 nIleLysGluTyrAspGluGlnPro...ThrTyrSerCysGlnLeuLeu 374
|||||
3666 ATGAGTCAAAAAATGATCAGGAATACAAATAATCAATGCAAAATG... 3711
|||||
374 heLysLysAspGluGluSerThrAspGluIleGlyLeuIleHis 390
|||||
3712 ..AATTTACAAGATAATAATGGGAATGATATAGGCTTTTATAGGATTCAT 3759
|||||
391 ArgPheTyrGluSerGlyIleValPheGluGluTyrLysAspTyrPheC 407
|||||
3760 CAGTTTAAATAATAGCT.....AAACTAGT 3785
|||||
407 sIleSerLysTyrTyrLeuLysGluValLysArgLysProTyrAsnLeu 424
|||||
3786 AGCAAGTATGTTATATAGACAAATAGAAAGATCT.....AGTAGGA 3829
|||||
```

alignment_scores:

```
424 ysLeuGlyCysAsnTrpGlnPheIleProLysAspGluGlyTrpThrGlu 440
|||||
3830 CTTTGGTTCCTCATGGGAATTTATCTCTAGATGATGATGGGAGAA 3879
|||||
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-915-136-27

seq_documentation_block:
; Sequence 27, Application US/08915136
; Patent No. 6290960
; GENERAL INFORMATION:
; APPLICANT: KINK, JOHN A.
; APPLICANT: THALLEY, BRUCE S.
; APPLICANT: PADHYE, NISHA V.
; APPLICANT: FIRCA, JOSEPH R.
; APPLICANT: STAFFORD, DOUGLAS C.
; TITLE OF INVENTION: VACCINE AND ANTITOXIN FOR TREATMENT AND
; PREVENTION OF C. DIFFICILE DISEASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MEDLEN & CARROLL, LLP
; STREET: 220 MONTGOMERY STREET, SUITE 2200
; CITY: SAN FRANCISCO
; STATE: CALIFORNIA
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 94104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/915,136
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/480,604
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/405,496
; FILING DATE: 16-MAR-1995
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/329,154
; FILING DATE: 25-OCT-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/161,907
; FILING DATE: 02-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/985,321
; FILING DATE: 04-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/429,791
; FILING DATE: 31-OCT-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: INGOLIA, DIANE E.
; REGISTRATION NUMBER: 40,027
; REFERENCE/DOCKET NUMBER: OPND-01763
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 705-8410
; TELEFAX: (415) 397-8338
; INFORMATION FOR SEQ ID NO: 27:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3891 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3888
; US-08-915-136-27
```

```

Quality: 831.50          Length: 450
Ratio: 2.607            Gaps: 10
Percent Similarity: 70.889 Percent Identity: 40.000

alignment_block:
US-09-910-186A-8 x US-08-915-136-27  ..

Align seg 1/1 to: US-08-915-136-27 from: 1 to: 3891

4 LysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeuArgTyrIly 20
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2602 CAATATATTAGAATATTATTAAATCTCTATATTGAATTTAAGATATGA 2651

20 sAspAsnLeuIleLeuAspLeuSerGlyTyrGlyAlaLysValGluVal 37
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2652 AAGTAATCATTTAATAGACTTTATCTAGGTATGCATCAAAAATAAATATTG 2701

37 yrAspGlyValCluLeuAsn.....AspLysAsnGlnPhelysLeuThr 51
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2702 GTAGTAAGATTAATTTTTGATCCAAATAGATAAAATCAAAATCAATTTAT 2751

52 SerSerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePh 68
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2752 AATTTAGAAAGTAGTAAAAATTCGAGCTAATTTAAAAAATGCTATTCTATA 2801

68 eAsnSerValPheLeuAspPheSerValSerPheThrIleArgIleProL 85
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2802 TAATAGTAGTATGAAAATTTTAGTACTAGCTTTTGGATGAAGAATTCCTA 2851

85 ySTyrLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIle 101
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2852 AGTATTTTAAAC.....AGTAAAGTCTAAATGAATGAATATACATA 2892

102 IleAsnCysMetLysAsnAsnSerGlyTyrIleSerIleArgGlyAs 118
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2893 ATAAATTCATGGAATAATTCAGATGGAAAGTATCACTTAATTAATGCG 2942

118 nArgIleIleThrLeuIleAspIleAsnGlyLysThrLysSerValP 135
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2943 TGAATAATATCTGGACTTTACAGGATACTCAGGAATAAAACAAGAGTAG 2992

135 hePheGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTTP 151
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
2993 TTTTAAATACAGTCAATGATTATATATATACAGATTATATAACACATGG 3042

152 PhePheValThrIleThrAsnAsn...LeuAsnAsnAlaLysIleTyr 167
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3043 ATTTTGTGAACATATCACATAATAGATTAAATAACTCTAAAAATTTATAT 3092

167 eAsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGluVal 184
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3093 AAATGGAGAGTAAATAGATCAAAAACCAATTTCAAAATTTAGGTAATATTC 3142

184 leAlaAsnGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThr 200
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3143 ATGCTAGTAATAATAATATGTTTAAATAGATGGTGTAGACATACACAT 3192

201 GlnPheIleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerG 217
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3193 AGATATATTGGATAAAATATTTTAATCTTTTGTATAAGGAATTAATATGA 3242

217 nSerAsnIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeu 234
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3243 AAAAGAAATCAAGAGTTTATATGATAATCAATCAAAATTCAGGTATTTAA 3292

234 ysAspPheTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPhe 250
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3293 AAGACTTTTGGGTGATTATTTACAATATGATAAACCATACATATATGTTA 3342

251 AsnAlaGlyAsnLysAsnSerTyrIleLysLeuLysLys..... 263
  ::::: ::::: ::::: ::::: ::::: ::::: ::::: ::::: :::::
3343 AATTTATATGATCCAAATAAATATGTCGATGTAAATAATGATAGGTATTAG 3392

```

```

264 .....AspSerProValGlyGluLeuLeuThrArgSerL 275
3393 AGGTTATATGTATCTTAAAGGGCCTAGAGTAGCGTAATGACTACAAACA 3442
275 ySTyRAsnGlnAsnSerLysTyRlleAsnTyRArgAspLeuTyRlleGly 291
3443 TTTATTATTAATTCAGT.....TTGTATATAGGGG 3471
292 GluLysPheIleleArgLysSerAsnSerGlnSerIleAsnAspAs 308
3472 ACAAAATTTATATAAAAAATGCTCTGGAAAT.....AAGATAA 3515
308 pIleValArgLysGluAspTyRlleTyRleuAspPheAsnLeuAsnG 325
3516 TATTGTTAGAAATAATGACGCTATATATATTAATAGTAGTTAAAAATA 3565
325 lngLutrpArgValTyRThrTyRLeuTyRPhelyLysGluGluGluLys 341
3566 AAGAATATATAGTTAGTACTAATGCAATCACAGCAGCGGTAGAAAAATA 3615
342 LeuPheLeuAlaProIleSerAspSerAspGluLeuTyRAsnThrIleG 358
3616 CTAAGTGCAATTAGAAATACCTGATGTAGGAAATCAAGTCAAGTAGTAGT 3665
358 nIleLysGluTyRAspGluClnPro...ThrTyRSerCysGlnLeuLeuP 374
3666 AATGAAGTCAAAAAATGATCAAGGAATAACAAATAAATGCAAAATG.... 3711
374 heLysLysAspGluGluSerThrAspGluIleGlyLeuIleGlyIleHis 390
3712 ..AATTACAGATAATAATGGAATGATAGCTTATAGGATTTCAT 3759
391 ArgPheTyRGlusergGlylleValPheGluGluTyRLeuLysAspTyRPhCy 407
3760 CAGTTTAATAATATAGCT.....AACTAGT 3785
407 sIleSerLysTrpTyRLeuLysGluValLysArgLysProTyRAsnLeuL 424
3786 AGCAAGTAATTGGTATAAATAGACAAATAGAAAGATCT.....AGTAGGA 3829
424 ysLeuGlyCysAsnTrpGlnPheIleProLysAspGluGlyTrpThrGlu 440
3830 CTTTGGGTGCTCATGGGAATTATTCTCTAGATGATGATGGGGAGAA 3879
seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-618-312A-1
seq_documentation_block:
; Sequence 1, Application US/07618312A
; Patent No. 5389540
; GENERAL INFORMATION:
; APPLICANT: Makoff Dr, Andrew J
; APPLICANT: Romanos Dr, Michael A
; APPLICANT: Clare Dr, Jeffrey J
; APPLICANT: Fairweather Dr, Neil F
; TITLE OF INVENTION: VACCINES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: 14th Floor
; STREET: 2200 Clarendon Boulevard,
; CITY: Arlington,
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22201
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/618.312A
; FILING DATE: 19910516
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:

```

```

; APPLICATION NUMBER: GB 8926832.0
; FILING DATE: 28-NOV-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9006097.1
; FILING DATE: 17-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Crawford Mc, Arthur R
; REGISTRATION NUMBER: 25,327
; REFERENCE/DOCKET NUMBER: 510-51
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 0101 703 8750400
; TELEFAX: 0101 703 5253468
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1359 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Clostridium tetani
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1356
; US-07-618-312A-1

```

alignment_scores:

```

Quality: 595.00 Length: 441
Ratio: 2.017 Gaps: 12
Percent Similarity: 66.893 Percent Identity: 32.200

```

```

alignment_block:
US-09-910-186a-8 x US-07-618-312A-1

```

```

Align seg 1/1 to: US-07-618-312A-1 from: 1 to: 1359

```

```

14 IleLeuAsnLeuArgTyrLysAspAsnLeuIleAspLeuSerGlyTy 30
|||||
67 ATTTAAATTTAGATATTAATGATATTAATATCGATATCTGGGT 116
|||||
30 rGlyAlaLysValGluValTyrAspGlyValGluLeu.....A 43
|||||
117 TAATTCATCTGTAATAACATATCCAGATGCTCAATTGGTCCCGGAATA 166
|||||
43 snAspLysAsnGlnPheLysLeuThrSerSerAlaAsnSerLysIleArg 59
|| |||
167 ATGGCAAA...GCAATACATTTAGTAAACAATGAATCTTCTGAAGTTATA 213
|||||
60 ValThrGlnAsnGlnAsnIlePheAsnSerValPheLeuAspPheSe 76
|||
214 GTGCATAAGCTATGATATGTAATGATATGATATGTTTAATAATTTTAC 263
|||||
76 rValSerPheTrpIleArgIleProLysTyrLysAsnAspGlyIleGlnA 93
|||||
264 CGTAGCTTTTGGTTGAGGTTCTCTAAAGTATCTGCTAGTCATTAGAAC 313
|||||
93 snTyrIleHisGlnGluTyrThrIleIleAsnCysMetLysAsnAsn... 108
|||||
314 AATATGGCAACAATGATAGTATCAATAATTAGCTCTATGAAAAACATAGT 363
|||||
109 .....SerGlyTrpLysIleSerIleArgGlyAsnArgIleIl 121
|||||
364 CTAATCAATAGGATCTGGTGGAGTGATCACTTAAAGGTAATAAATCAAT 413
|||||
121 etrpThrLeuIleAspIleAsnGlyLysThrLysSerValPheGluTr 138
|||||
414 ATGGACTTTAAAGATTCGCGGAGAGAGTAGACAAATAAATCTTTAGG. 462
|||||
138 yrAsnIleArgGluAspIleSerGluTyrIle...AsnArgTrpPhePhe 153
|||||

```

```

463 ..GATTTACCTGATAAATTTAATGCTTTATTAGCAAAATAAATGGTTTT 510
154 ValThrIleThrAsnAsn...LeuAsnAsnAlaLysIleTyrIleAsnGl 169
|||||
511 ATAACATTACTAATGATGATATCTTCTGCTAAATTTGTATATAATGG 560
|||||
169 yLysLeuGluSerAsnThrAspIleLysAspIleArgGluValIleAla 186
|||||
561 AGTACTTATCGGAAGTCGAGAAATTACTGTTAGGAGCTATTAGAGAGG 610
|||||
186 snGlyGluIleLlePheLysLeuAspGlyAspIleAspArgThrGlnPhe 202
|||||
611 ATAATAATATACATTAATAACTAGATAGATGTAATAATAATCAATAC 660
|||||
203 IleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGlnSerAs 219
|||||
661 GTTCTATTGATAAATTTAGGATATTTTGCAAAGCATTAATAATCCAAAGA 710
|||||
219 nIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLysLysasp 236
|||||
711 GATTGAAAAATTTATACACAGTTATTATCTATAACCTTTTAAAGAGACT 760
|||||
236 heTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPheAsnAla 252
|||||
761 TCTGGGAAACCCCTTACGATATGATACAGATATATTATTAAATACCAGA 810
|||||
253 GlyAsnLysAsnSerTyrIleLysLeuLysLysaspSerProValGlyGl 269
|||||
811 GCTTCTAGTTCTAAAGATGTTCAATTTGAAATAATATAACAGATTATATGTA 860
|||||
269 uIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleAsnTyrA 286
|||||
861 TTTGACAAATGCCCATCGTATACTAACGCGAAATTCGAATATATATATA 910
|||||
286 rgAspLeuTyrIleGlyGluLysPheIleIleArgLysSerAsnSer 302
|||||
911 GAAGGTTATATAATGGACTAAAAATTTATTATAAAAGATATACACCTAA 960
|||||
303 GlnSerIleAsnAspIleValArgLysGluAspTyrIleTyrIleuAs 319
|||||
961 AATGAATA...GATCTTTTGTAACTAGGTGATTTTATTAAATTA... 1005
|||||
319 pPhePheAsnLeuAsnGlnGluTrpArgValTyrThrTyr..... 332
|||||
1006 .TATGATCATATAACAATAATAGCAGCATTTGAGTTATCCGAAAGATG 1054
|||||
333 ..LysTyrPheLysLysGluGluLysLeuPheLeuAlaProIleSer 348
|||||
1055 GAAATGCCCTTTAATAATCTTTAGTAGAATTCCTAAGAGTAGGTTAATGCC 1104
|||||
349 AspSerAspGluLeuTyrAsnThrIleGlnIleLysGluTyrAspGluGl 365
|||||
1105 CAGGTATCCCTCTTTATAAAAAATGGAAGCAGTAAAAATTCGCTGATTT 1154
|||||
365 nProThrTyrSerCysGlnLeuLeuPheLysLysAspGluLysSerThrA 382
|||||
1155 AAAAACTATTCTGTACAACTTAAATTTATGATGATAAAAAATGCATCT. 1203
|||||
382 spCluIleGlyLeuIleGlyIleHisArgPheTyrGluSerGlyIleVal 398
|||||
1204 ....TAGGACTAGTAGGTACCCAT.....AATGTCAAATAGGC 1239
|||||
399 PheGluGluTyrLysAspTyrPheCysIleSerLysTrpTyrIleuLysGl 415
|||||
1240 AAGCATCCAAATAGGATATATTAAATTCGAAGCACTGGTACTTTAATCA 1289
|||||
415 uValLysArgLysProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheI 432
|||||
1290 TTTAAAGATAAA.....ATTTTAGCATGTGATTGCTACTTTCG 1327
|||||
432 leProLysaspGluGlyTrpThr 439
|||||
1328 TACCTACAGATGAAGGATGGACA 1350

```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-280-228-1

```
seq_documentation_block:
; Sequence 1, Application US/08280228
; Patent No. 5571694
; GENERAL INFORMATION:
; APPLICANT: Makoff Dr, Andrew J
; APPLICANT: Romanos Dr, Michael A
; APPLICANT: Clare Dr, Jeffrey J
; APPLICANT: Fairweather Dr, Neil F
; TITLE OF INVENTION: VACCINES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NIXON & VANDERHYE P.C.
; STREET: 1100 NO. 5571694th Glebe Road
; CITY: Arlington,
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/280,228
; FILING DATE: 25-JUL-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/618,312
; FILING DATE: 27-NOV-1990
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 8926832.0
; FILING DATE: 28-NOV-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9006097.1
; FILING DATE: 17-MAR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Wilson, Mary J.
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 117-163
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1359 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ORIGINAL SOURCE:
; ORGANISM: Clostridium tetani
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1356
; US-08-280-228-1
```

alignment_scores:
Quality: 595.00 Length: 441
Ratio: 2.017 Gaps: 12
Percent Similarity: 66.893 Percent Identity: 32.200

alignment_block:
US-09-910-186a-8 x US-08-280-228-1 ..

Align seg 1/1 to: US-08-280-228-1 from: 1 to: 1359

```
14 IleLeuAsnLeuArgTyrLysAspAsnAsnLeuIleAspLeuSerGlyTyr 30
|||||
67 ATTTTAAATTTAGATATTAATAATGATTAATAATCGATATATCGGGT 116
|||||
30 rGlyAlaLysValGluValTyrAspGlyValGluLeu.....A 43
:|||||
117 TAATTCATCTCTAATAACATATCCAGATGCTCAATTTGTCGCCGGAATAA 166
|||||
43 snAspLysAsnGlnPheLysLeuThrSerSerAlaAsnSerLysIleArg 59
|||||
167 ATGCCAAA...GCAATACATTTAGTAACAATGAATCTCTCGAAGTTATA 213
|||||
60 ValThrGlnAsnGlnAsnIlePheAsnSerValPheLeuAspPheSe 76
|||||
214 GTGCATAAAGCTATGGATATTCATATAATGATATGTTTAAATAATTAC 263
|||||
76 rValSerPheTrpIleArgIleProLysTyrLysAsnAspGlyIleGlnA 93
:|||||
264 CGTTAGCTTTTGGTTGAGGTTCTCTAAAGTATCTGCTAGTCATTAGAAC 313
|||||
93 snTyrIleHisAsnGluTyrThrIleIleAsnCysMetLysAsnAsn... 108
:|||||
314 AATATGGCACAATAGATGATTCATAATAGCTCTATGAAAAACATAGT 363
|||||
109 .....SerGlyTrpLysIleSerIleArgGlyAsnArgIleI 121
|||||
364 CTATCAATAGGATCTGGTTGGAGTGATCACTTAAAGGTAATAACTTAAT 413
|||||
121 eTrpThrLeuIleAspIleAsnGlyLysThrLysSerValPhePheGluT 138
:|||||
414 ATGGACTTTAAAGATTCGCGGGAGGAAGTTAGACAAATAACTTTTAGG. 462
|||||
138 yrAsnIleArgGluAspIleSerGluTyrIle...AsnArgTrpPhePhe 153
:|||||
463 ..GATTACCTGATAAAATTAATGCTTATTTAGCAAAATAAATGGGTTTT 510
|||||
154 ValThrIleThrAsnAsn...LeuAsnAsnAlaLysIleTyrIleAsnGl 169
:|||||
511 ATAATAATTAATAGATAGATTAATCTCTGCTAATTTGTATATAAATGG 560
|||||
169 YLysLeuGluSerAsnThrAspIleLysAspIleArgGluValIleAlaA 186
:|||||
561 AGTACTTATGGAGAGTCAGAAATTAAGTCTAGGAGCTATTAGAGAGG 610
|||||
186 snGlyGluIlePheLysLeuAspGlyAspIleAspArgThrGlnPhe 202
:|||||
611 ATAATAATTAACATTAACACTAGATAGATGTAATAATAATAATCAATAC 660
|||||
203 IleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGlnSerAs 219
:|||||
661 GTTCTCTATTGATAAATTTAGGATATTTTGCAAGCATTAATCCAAAGA 710
|||||
219 nIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeuLysAspP 236
:|||||
711 GATTGAAAAATTAACACAAAGTTATTTATCTATAACCTTTTAAAGAGACT 760
|||||
236 heTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPheAsnAla 252
:|||||
761 TCTGGGGAACCCCTTACGATATGATACAGAAATATTATTATACCAGTA 810
|||||
253 GlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProValGlyGl 269
:|||||
811 GCTTCTAGTTCTAAAGATGTTCAATTGAAAAATATACACAGATTATATGA 860
|||||
269 uIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleAsnTyrA 286
:|||||
861 TTTGACAAATCGCCCATCGTATACTACTAACGGAAATTTGAATATATATA 910
|||||
286 rGAspLeuTyrIleGlyGluLysPheIleIleArgArgLysSerAsnSer 302
:|||||
911 GAGGTTATATAATGGACTAAATTTATTATAAAAAGATATACACCTAAT 960
|||||
303 GlnSerIleAsnAspAspIleValArgLysGluAspTyrIleTyrLeuAs 319
```



```
561 CGTTCTGATGGCTCCGCTGAAATCACTGGTCTGGCGCTATCCGTGAGG 610
186 snGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThrGlnPhe 202
611 ACACCAACATCACTCTTAGCTGGACCGTTGGCAACAACCAACCAAGTAC 660
203 IleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGlnSerAs 219
661 GTATCCATCGACAAGTTCGGTATCTCTGCAAGACACTGAACCCGAAGA 710
219 nIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeuLysAspP 236
711 GATCGAAAACCTGTATACCAAGCTAGCTGTCTATCACTTCCTCGCTG 760
236 heTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPheAsnAla 252
761 TCTGGGGTAACCGCTGGCTTACGACACGAATATTACTGTATCCGGTGA 810
253 GlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProValGlyG 269
811 GCTTCTAGCTTAAGACGTTCAAGCTGAACATCACTGACTACATGTA 860
269 uIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleAsnTyrA 286
861 CCGTACCAACGCGCGTCTCTACACTAAGGTAAGCTGAACATCTACTACC 910
286 rGAspLeuTyrIleGlyGluLysPheIleIleArgLysSerAsnSer 302
911 GAGCTGTGTACAAACCGCTGAATTCATCACTCAACGCTACACTCCGAAC 960
303 GlnSerIleAsnAspAspIleValArgLysGluAspTyrIleTyrLeuAs 319
961 AACGAAATC...GATTCITTCGTTAAATCTGGTACTTCATCAAACTG... 1005
319 pPhePheAsnLeuAsnGlnGluTyrPArgValTyrThrTyr..... 332
1006 .TACGTTTCTTACAAACAACAACGACATCGTTGGTTACCGGAAAGAG 1054
333 ..LysTyrPheLysLysGluGluGluLysLeuPheLeuAlaProIleSer 348
1055 GTAACGCTTCAACACCTGGACAGAAATCTCGGTGGTTGATTACACGCT 1104
349 AspSerAspGluLysTyrAsnThrIleGlnIleLysGluTyrAspGluG 365
1105 CGGGTATCCCGCTGTACAAAAAATGGAAGCTGTAAACTGCGTGACCT 1154
365 nProThrTyrSerCysGlnLeuLeuPheLysLysAspGluGluSerThrA 382
1155 GAAAACCTACTCTGTACGCTGAAACTGTACGACGACAAAAACGCTTCT. 1203
382 spGluIleGlyLeuIleGlyIleHisArgPheTyrGluSerGlyIleVal 398
1204 .....CTGGTCTGGTGGTACCCAC.....AACGGTCAGATCGGT 1239
399 PheGluGluTyrLysAspTyrPheCysIleSerLysTyrTyrLeuLysG 415
1240 AACGACCGCAACCGTGACATCTGTACGCTTCTAACTGCTACTTCAACCA 1289
415 uValLysArgLysProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheI 432
1290 CCTGAAGACAAA.....ATCTGGGTTGCGAGCTGGTACTTCG 1327
432 leProLysAspGluGlyTrpThr 439
1328 TTCCGACCGATGAAGTTGGACC 1350
```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-280-228-3

seq_documentation_block:
; Sequence 3, Application US/08280228
; Patent No. 5571694
; GENERAL INFORMATION:
; APPLICANT: Makoff Dr, Andrew J

```
; APPLICANT: Romanos Dr, Michael A  
; APPLICANT: Clare Dr, Jeffrey J  
; APPLICANT: Fairweather Dr, Neil F  
; TITLE OF INVENTION: VACCINES  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: NIXON & VANDERHVE P.C.  
; STREET: 1100 No. 5571694th Glebe Road  
; CITY: Arlington,  
; STATE: Virginia  
; COUNTRY: U.S.A.  
; ZIP: 22201-4714  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/280,228  
; FILING DATE: 25-JUL-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 8926832.0  
; FILING DATE: 28-NOV-1989  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9006097.1  
; FILING DATE: 17-MAR-1990  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Wilson, Mary J.  
; REGISTRATION NUMBER: 32,955  
; REFERENCE/DOCKET NUMBER: 117-163  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 816-4000  
; TELEFAX: (703) 816-4100  
; TELEX: 200797 NIXN UR  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1359 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: cDNA  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Clostridium tetani  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 1..1356  
; US-08-280-228-3  
  
alignment_scores:  
Quality: 594.00 Length: 441  
Ratio: 2.014 Gaps: 12  
Percent Similarity: 66.893 Percent Identity: 31.973  
  
alignment_block:  
US-09-910-186A-8 x US-08-280-228-3 ..  
  
Align seg 1/1 to: US-08-280-228-3 from: 1 to: 1359  
  
14 IleLeuAsnLeuArgTyrLysAspAsnLeuIleAspLeuSerGlyTy 30  
67 ATTCTGAACCTTGACATCAACAACGATATATCTCCGACATCTCGTTT 116  
30 rGlyAlaLysValGluValTyrAspGlyValGluLeu.....A 43  
117 CAATCTCTCTGTATCATCATATCCAGATGCTCAATTGGTGGCGGCATCA 166
```

```
43  snAspLysAsnGlnPheLysLeuThrSerSerAlaAsnSerLysIleArg 59
||| |||
167  ACGCAAA...GCTATCCACCTGGTTAAACAACGAATCTCTGAAGTATC 213
||| |||
60  ValThrGlnAsnGlnAsnIleIlePheAsnSerValPheLeuAspPheSe 76
||| |||
214  GTGCACAAGCCCATGGACATCGAATACAACGACATGTTCACAACATTCAC 263
||| |||
76  rValSerPheTrpIleArgIleProLysTyrLysAsnAspGlyIleGlnA 93
||| |||
264  CGTAGCTTCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCTGGCT 313
||| |||
93  snTyrIleHisAsnGluTyrThrIleIleAsnCysMetLysAsnAsn... 108
||| |||
314  AGTACGGCAGCTAAGAGTACTCATCATCAGCTCATATGAAGAACAACATCC 363
||| |||
109  .....SerGlyTrpLysIleSerIleArgGlyAsnArgIleI 121
||| |||
364  CTGTCCATCGGCTCTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 413
||| |||
121  eTrpThrLeuIleAspIleAsnGlyThrLysSerValPhePheGluTr 138
||| |||
414  CTGGACTCTGAAGAGTCCCGCGGCGAAGTCTGTGATCAGTCACTTCCGC. 462
||| |||
138  yRAsnIleArgGluAspIleSerGluTyrIle...AsnArgTrpPhePhe 153
||| |||
463  ..GACCTGCGGACAGTTCACCGGTACCTGGCTGACCAATGGGTTTC 510
||| |||
154  ValThrIleThrAsnAsn...LeuAsnAsnAlaLysIleTyrIleAsnG 169
||| |||
511  ATCATTATCATCAACGATCTGTCTGTCTGTCTGTCTGTCTGTCTGTCTGT 560
||| |||
169  yLysLeuGluSerAsnThrAspIleLysAspIleArgGluValIleAla 186
||| |||
561  CGTCTCTGATGGGCTCGCTGAAATCACTGGTCTGGCGCTATCCGTCAGG 610
||| |||
186  snGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThrGlnPhe 202
||| |||
611  ACAACACATCATCTTAAGCTGGACCGTGTGCAACACACACACACACAGTAC 660
||| |||
203  IleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGlnSerAs 219
||| |||
661  GTATCATCATGACAAAGTCCGTATCTTCTGCAAAAGCACTGAACCCGAAAG 710
||| |||
219  nIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLysLeuAsp 236
||| |||
711  GATCGAAAACCTATATACCACTACCTGTCTATCACTCTCTCTCGTGACT 760
||| |||
236  heTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPheAsnAla 252
||| |||
761  TCTGGGGTACCCGCGGTGACGACCGCAATATTAATCTGTATCCCGGTA 810
||| |||
253  GlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProValGlyG 269
||| |||
811  GCCTTAGCTTAAGACGCTTACGCTGCTGCTGCTGCTGCTGCTGCTGCTG 860
||| |||
269  uIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleAsnTyrA 286
||| |||
861  CTTGACCAACGCGCTCTCTACTAAGCGTAAACATCACTACTACTACTACC 910
||| |||
286  rGAspLeuTyrIleGlyGluLysPheIleIleArgLysSerAsnSer 302
||| |||
911  GAGCTGTGACAAAGCGCTGAAATCATCATCAACGCTACATCCCGAAC 960
||| |||
303  GlnSerIleAsnAspIleValArgLysGluAspTyrIleTyrLeuAs 319
||| |||
961  AACGAAATC...GATCTTTCGTTAAATCTGGTGACTTCATCAAACTG... 1005
||| |||
319  pPheAsnLeuAsnGlnGluTrpArgValTyrThrTyr..... 332
||| |||
1006  .TACGTTTCTTACAACAACAACAACACATCGTTGGTTACCCCGAAGACG 1054
||| |||
```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-668-381A-6

seq_documentation_block:

; Sequence 6, Application US/08668381A

; Patent No. 5780024

; GENERAL INFORMATION:

; APPLICANT: Brown, Robert H.

; APPLICANT: Fishman, Paul S.

; APPLICANT: Francis, Jonathan W.

; APPLICANT: Hosler, Betsy A.

; TITLE OF INVENTION: SUPEROXIDE DISMUTASE/TETANUS TOXIN

; TITLE OF INVENTION: FRAGMENT C HYBRID PROTEIN

; NUMBER OF SEQUENCES: 6

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Fish & Richardson P.C.

; STREET: 225 Franklin Street

; CITY: Boston

; STATE: MA

; COUNTRY: USA

; ZIP: 02110-2804

; COMPUTER READABLE FORM:

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patent In Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/668,381A

; FILING DATE: 21-JUN-1996

; CLASSIFICATION: 514

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/000,473

; FILING DATE: 23-JUN-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Clark, Paul T.

; REGISTRATION NUMBER: 30,164

; REFERENCE/DOCKET NUMBER: 00786/269001

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 617/542-5070

; TELEFAX: 617/542-8906

; TELEX: 200154

; INFORMATION FOR SEQ ID NO: 6:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 1858 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

```

seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-08-110-786A-7
seq_documentation_block:
; Sequence 7, Application US/08110786A
; Patent No. 5443966
; GENERAL INFORMATION:
; APPLICANT: FAIRWEATHER, Neil Fraser
; APPLICANT: MAKOFF, Andrew Joseph
; TITLE OF INVENTION: Expression of tetanus toxin fragment C
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon & Vanderhye P.C.
; STREET: 1100 No. 5443966th Glebe Road
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22201-4714
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:

```

```

; APPLICATION NUMBER: US/08/110,786A
; FILING DATE: 23-AUG-1993 1991
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: US 07/777,337
; FILING DATE: 29-NOV-1991
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB90/00943
; FILING DATE: 20-JUN-1990
; APPLICATION NUMBER: GB 8914122.0
; FILING DATE: 20 June 1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Mary J. Wilson
; REGISTRATION NUMBER: 32,955
; REFERENCE/DOCKET NUMBER: 117-134
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 816-4000
; TELEFAX: (703) 816-4100
; TELEX: 200797 NIXN UR
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1359 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA
; ORIGINAL SOURCE:
; ORGANISM: Clostridium tetani
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..1359
; US-08-110-786A-7

```

```

alignment_scores:
  Quality: 575.00      Length: 441
  Ratio: 1.969         Gaps: 12
  Percent Similarity: 66.213  Percent Identity: 31.519

```

```

alignment_block:
  US-09-910-186A-8 x US-08-110-786A-7

```

```

Align seg 1/1 to: US-08-110-786A-7 from: 1 to: 1359

```

```

14 1leLeuAsnLeuArgTyrLysAspAsnAsnLeuLeuLeuLeuSerGlyTy 30
|||||
67 ATTCTGAACCTGGACATCAACAACGATATTATCTCCGACATCTCTGGTTC 116
30 rGlyAlaLysValGluValTyrAspGlyValGluLeu.....A 43
|||||
117 CAATCTCTAGGATCAGATATCCAGATGCTCAATGGTGCCCGAATAA 166
43 snAspLysAsnGlnPheLysLeuThrSerSerAlaAsnSerLysIleArg 59
|||||
167 ATGGCAAA...GCAATACATTAGTAACAATGAATCTCTGGAAGTTATA 213
60 ValThrGlnAsnGlnAsnIleIlePheAsnSerValPheLeuAspPheSe 76
|||||
214 GTGCATAAAGCATGGATATTGAATATAATGATATGTTTAATAATTTTAC 263
76 rValSerPheTrpIleArgIleProLysTyrLysAsnAspGlyIleGlnA 93
|||||
264 CGTAGCTTTGGTGGAGGTTCTTAAGTATCTAGTATCTATTAGAAC 313
93 snTyrIleHisAsnGluTyrThrIleLeuAsnCysMetLysAsnAsn... 108
|||||
314 AATATGGCAAAATGAGTATTCAATAAATAGCTCTATGAAAAAACAATAGT 363
109 .....SerGlyTrpLysIleSerIleArgGlyAsnArgIleI 121
|||||
364 CTATCAATAGGATCTGGTTGGAGTGATCATCTTAAAGGTAATAACTTAAT 413

```

```

121 eTrpThrLeuIleAspIleAsnGlyLysThrLysSerValPhePheGluT 138
|||||
414 ATGGACITTAAGAAGATTCGCGGGAGAGTTAGACAAATAACTTTTAGS. 462
138 yrAsnIleArgGluAspIleSerGluTyrIle...AsnArgTrpPhePhe 153
|||||
463 ..GATTTACCTGATAAATTTAATGCTTATTAGCAAAATAAATGGTTT 510
154 ValThrIleThrAsnAsn...LeuAsnAsnAlaLysIleTyrIleAsnG 169
|||||
511 ATAATACTATTACTAATGATAGATATCTCTGCTAATTTGATATATAAAT 560
169 yLysLeuGluSerAsnThrAspIleLysAspIleArgGluValIleAla 186
|||||
561 AGTACTTATGGAACTGCAGAAATTTACTGGTTTAGGACCTATTACAGAG 610
186 snGlyGluIlePheLysLeuAspGlyAspIleAspArgThrGlnPhe 202
|||||
611 ATAATAATATAACATTTAAACTAGATAGTAAATAAATAAATAATCAATAC 660
203 IleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGlnSerAs 219
|||||
661 GTTCTTATTGATAAATTTAGGATATTTTGCAAAGCATTTAAATCCAAAGA 710
219 nIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeuLysAsp 236
|||||
711 GATTGAAAAATTTATACACAAGTTATTTATCTATTAACCTTTTAAAGACT 760
236 heTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPheAsnAla 252
|||||
761 TCTGGGAAACCTTTACGATATGATACAGAATATATTATTAATACCAGTA 810
253 GlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProValGlyCl 269
|||||
811 GCTTCTAGTCTTAAAGATGTTCAATTTGAAATAATATAACAGATTATATGA 860
269 uIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleAsnTyrA 286
|||||
861 TTTGACAAATCGCCATCGTATACTACGGAAATTTGAATATATATATATA 910
286 rgAspLeuTyrIleGlyLysPheIleIleArgArgLysSerAsnSer 302
|||||
911 CAAGTTATATAATGGACTAAATTTATATAAAGATATACACCTAAT 960
303 GlnSerIleAsnAspAspIleValArgLysGluAspTyrIleTyrLeuAs 319
|||||
961 AATGAAATA...GATCTTTTAAATCAGGTGATTTTATTAATAA... 1005
319 pPhePheAsnLeuAsnGlnGluTrpArgValTyrThrTyr..... 332
1006 .TATGTATCATATAACAATAATAGACACATTTGAGTTATCCGAAGATG 1054
333 ..LysTyrPheLysLysGluGluLysLeuPheLeuAlaProIleSer 348
|||||
1055 GAAATGCTTTAATAATCTTATAGAAATTTAAGAGTAGGTATATAATGCC 1104
349 AspSerAspGluLeuTyrAsnThrIleGlnIleLysGluTyrAspGluCl 365
|||||
1105 CCAGGTATCCCTCTTTATAAAAAATGGAAGCAGTAAATTTGCGTGATTT 1154
365 nProThrTyrSerCysGlnLeuLeuPheLysLysAspGluGluSerThra 382
|||||
1155 AAAAACTATTTCTGACAACTTAAATATATATGATGATGATAAATGCATCT. 1203
382 spGluIleGlyLeuIleGlyIleHisArgPheTyrGluSerGlyIleVal 398
|||||
1204 ....TTAGGACTAGTAGTACCCAT.....AATGGTCAAAATAGGC 1239
399 PheGluGluTyrLysAspTyrPheCysIleSerLysTyrTyrLeuLysG 415
|||||
1240 AACGATCCAAATAGGATATATTAATTCGAAGCAACGTGCTACTTTAATCA 1289
415 uValLysArgLysProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheI 432

```


OM of: US-09-910-186a-8 to: EST:* out_format : pfs

Date: Sep 2, 2002 3:28 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=framet_p2n.model -DEV=xlh  
-O=/cgn2_1/USPPO_spool/US09910186/runat_29082002_134725_15979/app_query.fasta_1.505  
-DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -QGAPOP=4.500  
-GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-DELEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -NATRIX=blomsum62 -TRANS=human40.cdi  
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -HEAPSIZE=500  
-MINLEN=0 -MAXLEN=200000000 -USER=US09910186_@CGL1_1_2803  
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30  
-NO_XLPXY -WAIT -THREADS=1
```

Search information block:

Query: US-09-910-186a-8

Query length: 440

Database: EST:*

Database sequences: 13736207

Database length: 1841457050

Search time (sec): 1596.540000

score_list:

Sequence	Strd	Orig	Zscore	EScore	Len	Documentation
gb_est2:BM273933	+	129.50	241.45	0.0003	587	BM273933 PfESToa61e08.y1 Plasm
gb_est2:BM276542	+	128.00	239.11	0.0005	550	BM276542 PfESToa64g04.y1 Plasm
gb_est2:BM275974	+	128.00	238.95	0.0005	559	BM275974 PfESToa68e05.y1 Plasm
gb_est2:BM275232	+	128.00	238.93	0.0005	560	BM275232 PfESToa65h03.y1 Plasm
gb_est2:BM170629	+	123.00	228.87	0.0017	565	BM170629 EST573152 PyBS Plasmod
gb_est2:BM164577	+	123.00	224.81	0.0028	628	BM164577 EST567100 PyBS Plasmod
gb_est1:AL557249	+	114.00	205.54	0.0334	963	AL557249 LTI_FL012_TCL
gb_gss:BM150480	+	113.00	204.15	0.0400	907	BM150480 ENQB957FB Entamoeba H
gb_est2:BM276398	+	112.50	210.47	0.0178	439	BM276398 PfESToa82f02.y1 Plasm
gb_est1:AL557255	+	112.50	202.69	0.0482	950	AL557255 LTI_FL012_TCL
gb_gss:AL524527	+	112.00	205.64	0.0330	642	AL524527 233pbc08 Pb MBN #21 Pl
gb_est2:BG603391	+	111.50	205.28	0.0437	603	BG603391 EST502481 Plasmodium Y
gb_gss:BA53173	-	111.50	203.45	0.0437	723	BA53173 BOKB67TR BOKB Brassic
gb_est2:BM163086	+	111.00	201.28	0.0578	812	BM163086 EST565609 PyBS Plasmod
gb_gss:AL549268	-	110.50	199.89	0.0690	844	AL549268 ENTD208TF Entamoeba hi
gb_est2:BM167654	+	109.00	197.25	0.0968	815	BM167654 EST562778 PyBS Plasmod
gb_gss:BM160255	-	108.50	197.66	0.0919	709	BM160255 EST562778 PyBS Plasmod
gb_gss:BM158617	-	107.00	192.54	0.1773	876	BM158617 ENTE16TR Entamoeba hi
gb_gss:AL523332	-	107.00	192.15	0.1862	910	AL523332 ENTCG60TF Entamoeba hi
gb_gss:AL523372	-	107.00	192.15	0.1862	910	AL523372 ENTCX84TF Entamoeba hi
gb_est2:BM164327	+	106.50	192.43	0.1797	802	BM164327 EST566850 PyBS Plasmod
gb_gss:BM159101	+	106.50	191.24	0.2092	902	BM159101 ENTRW22TF Entamoeba hi
gb_gss:CN507DVE	+	106.50	190.89	0.2189	934	AL440976 T3 end of clone BDOAAC
gb_est2:BM1815339	+	106.00	196.00	0.1137	510	BM1815339 PfESToa61e03.y1 Plasm
gb_est2:BM168385	+	106.00	192.26	0.1838	739	BM168385 EST570908 PyBS Plasmod
gb_gss:CN5061KD	-	106.00	190.05	0.2440	920	AL404291 T7 end of clone AT0AAC
gb_gss:AL689211	+	106.00	189.72	0.2543	950	AL689211 ENTRILTR Entamoeba hi
gb_est2:BM168450	+	105.50	191.30	0.2077	736	BM168450 EST570973 PyBS Plasmod
gb_est2:BM159176	+	105.50	191.04	0.2147	755	BM159176 EST561699 PyBS Plasmod
gb_est2:BM169979	+	105.50	190.99	0.2162	759	BM169979 EST572502 PyBS Plasmod
gb_gss:AL523106	-	105.50	189.59	0.2587	872	AL523106 ENTCJ29TR Entamoeba hi
gb_gss:CN506161	-	105.50	189.41	0.2648	888	AL339887 T7 end of clone AS0AAC
gb_est1:AW688370	+	105.00	191.13	0.2123	678	AW688370 NF006603S3RIF1000 Devel
gb_est2:BM162427	+	104.50	189.87	0.2496	696	BM162427 EST564950 PyBS Plasmod
gb_gss:AO449733	-	104.50	189.25	0.2702	740	AO449733 500003H03.x2 CpiOWAM13
gb_gss:BM150734	+	104.50	187.55	0.3361	876	BM150734 ENTPF74TF Entamoeba hi
gb_gss:AL5235737	+	104.50	187.24	0.3495	903	AL5235737 ENTB51TR Entamoeba hi
gb_gss:CN50638V	+	104.50	185.81	0.4201	1041	AL401285 T3 end of clone AS0AAC
gb_htc:AK003217	+	104.00	181.60	0.7206	1580	AK003217 Mus musculus 18 days
gb_est2:BM169075	+	104.00	189.23	0.2711	672	BM169075 EST571598 PyBS Plasmod

gb_gss:BM136854 + 104.00 186.69 0.3752 864 ! BM136854 ENT0032TR Entamoeba
gb_gss:AL5234894 + 104.00 186.57 0.3808 874 ! AL5234894 ENTRU16TR Entamoeba
gb_est2:BM274388 + 104.00 172.00 2.476 3707 ! C83838 C83838 Dictyostellium
gb_est2:BM273951 + 103.00 189.12 0.2746 557 ! BM274388 PfESToa61e08.y1 Pl
gb_est2:BM273933 + 103.00 188.66 0.2913 583 ! BM273951 PfESToa61e08.y1 Pl

seq_name: gb_est2:BM273933

seq_documentation_block:

LOCUS BM273933 587 bp mRNA linear EST 20-DEC-2001
DEFINITION PfESToa61e08.y1 Plasmodium falciparum 3D7 gametocyte cDNA library
Plasmodium falciparum 3D7 cDNA 5', mRNA sequence.

ACCESSION BM273933

VERSION BM273933.1 GI:17967234

KEYWORDS EST.

SOURCE Plasmodium falciparum 3D7.

ORGANISM Plasmodium falciparum 3D7

REFERENCE 1 (bases 1 to 587)

AUTHORS Tang, K., Cole, R., Chakrabarti, D., Haywood, R., Clifton, S., Pape, D.,

Marra, M., Hillier, L., Martin, J., Wylie, T., Dente, M., Theising, B.,

Bowers, Y., Gibbons, M., Ritter, E., Bennett, J., Jentes, E., Ronko, I.,

Tsagarelshvili, R., Belaygorod, L., Franklin, C., Carr, L., Grow, A.,

Maguire, L., Richey, J., Wadkins, J., Kennedy, S., Levinso, D.,

Waterston, R., Wilson, R. and Sibley, D.

WashU Plasmodium EST Project

Unpublished (2001)

CONTACT: L. David Sibley

WashU Plasmodium EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800

Fax: 314 286 1810

Email: est@watson.wustl.edu

Library was constructed by R. Haywood. DNA sequencing by:

Washington University Genome Sequencing Center For information on

obtaining a clone please contact: L. David Sibley

(sibley@borcim.wustl.edu), Washington University

Seq primer: -40UP from Gibco

High quality sequence stop: 424.

FEATURES

source

1..587

/organism="Plasmodium falciparum 3D7"

/db_xref="taxon:36329"

/clone_lib="Plasmodium falciparum 3D7 gametocyte cDNA

library."

/dev_stage="gametocyte (stage III-V)"

/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"

/note="Vector: pBluescript SK plus; Site 1: EcoRI; Site 2:

XhoI. The library was constructed by R Haywood. cDNAs were

synthesized from gametocyte poly(A)+ RNA by oligo d(T)

priming, size-selected and directionally cloned into the

EcoRI (5' end) to XhoI (3' end) sites of the Uni-ZAP XR

lambda vector (Stratagene). The primary library was mass

excised as phagemid using the ExAssist helper phage

(Stratagene). Clones were mass excised using the ExAssist

helper phage (Stratagene), the phagemids were precipitated

with PEG 8000 and extracted with phenol/chloroform.

Phagemid DNA was electroporated into DH10B cells. Clone

Availability: David Sibley, Washington University."

alignment_scores:

Quality: 129.50 Length: 208

Ratio: 1.136 Gaps: 9

Percent Similarity: 54.808 Percent Identity: 24.519

alignment_block:

US-09-910-186A-8 x BM273933 ..

Align seg 1/1 to: BM273933 from: 1 to: 587

TITLE


```

207 TyrPhe..... 208
    |||:::
139 TATTATGATACCTGCATAAAATAATATATAGTAAGGAACAACACAGTTGA 188
    ..|||:::
209 SerIlePheAnThrGIuLeuSerGlnSerAsnIleGluGluArgTyrL 225
    :|||:::|||||:::|||||:::
189 TAATATATATAGAACCAATATGATAGCTCTCCAAAAAATATATAGAATA 238
    :|||:::|||||:::|||||:::
225 ysaiEclnSerTyrSerGluTyrLeu.....Lysasp 235
    :|||:::|||||:::|||||:::
239 ATATGAATTCATTGGTAGATACCATAAACAATATTATAAAGAATAAT 288
    :|||:::|||||:::|||||:::
236 PheTrpGly.....AsnProLeuMetTyrAsnLySgluTyrTyrMetPh 250
    :|||:::|||||:::|||||:::
289 ATTATGGTGACAAATAAATAATATAAATTAATAATTCACAAATTCAAAT 338
    :|||:::|||||:::|||||:::
250 easnAlGlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProv 267
    :|||:::|||||:::|||||:::
339 AAATTGTAATGAAAANAACGTTCATCAATAATAAATAAAGT..... 381
    :|||:::|||||:::|||||:::
267 alGlGluIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIle 283
    |||||:::|||||:::|||||:::
382 ..GGCTTTATAAAAACATATCCTTTTAATATATAAAAAATCTTATGCC 429
    :|||:::|||||:::|||||:::
284 AsnTyrArgAspLeuTyrIleGlyGluLysPheIleIleArgArgLysse 300
    :|||:::|||||:::|||||:::
430 ACATATAAACAG.....AAACAATAATAATTAATAAATAAATC 467
    :|||:::|||||:::|||||:::
300 rAsnSerGlnSerIleAsnAspAspIleValArgLysGluAspTyriler 317
    |||||:::|||||:::|||||:::
468 ACTTATACAGAAAANAAGGGGAAAAC.....AATTATATAT 502
    :|||:::|||||:::|||||:::

```

```

430 ACAATATGACAG.....AACCAATATGATTAATGACATGATC 407
300 rASnSerGIleASnAspIleValArgLysGluAspTyrIleIet 317
      |||:::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
468 ACTATACAGAAAGGGGAAAC.....AATTATATAT 502

317 yr 317
      ||
503 AT 504

```

Fraser, C.M. and Carucci, D.J.
Plasmodium yoelii EST project at TIGR
Unpublished (2001)
Contact: Jane Carlton
Parasite Genomics Group
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301-530-9319
Fax: 301-838-0208
Email: carlton@tigr.org
For clone info, please contact the Malaria Research and Reference
Reagent Resource Center, ATCC
<http://www.malaria.mr4.org/mr4pages/index.html>
Seq primer: ADP.

	Location/Qualifiers
1. .565	
/organism="Plasmodium yoelii"	
/strain="17XL"	
/db_xref="taxon:73239"	
/clone="PYCQ195"	
/clone_lib="PyBS"	
/dev_stage="Asexual blood stages"	
/lab_host="E. coli XL-1 Blue"	

[illegible]

```
||||| ||| ||| ::||| ::||| ::||| ::||| ::|||
475 ...CTCAAAATGTTT...CGTAATGGGATGATGATGAAGAGATAAC... 516
248 rMetPheAsnAlaGlyAsnLysAsn.....SerTyrIleL 260
||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
517 .....AGGGTCAGGATCAGTGAAGCATTTGGCAGATTACATCA 556
260 ysLeuTyrLysAspSerProValGlyGluLe..... 270
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
557 GGCAACAAAAAGTGAACCCATTCAAGAAATTCGGGACTTAGCAGAAATC 606
271 .....LeuThrArgSerLys.....TyrAsnGlnAsnSe 280
||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
607 ACCACTCTTGATCGCACCAAGAAATATCATTTGGATATTTTGACGAAAA 656
280 rLysTyrIleAsnTyrArgAspLeuTyrIleGlyGluLysPheIleLea 297
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
657 GGACTCGGACCACTATAGA.....GTTTPTG 682
297 rgArgLysSerAsnSerGlnSerIleAsnAspAlleValArgLysGlu 313
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
683 AACGAGTAGCGAAT.....ATTTCATGAT 708
314 AspTyrIleTyrLeuAspPheAsnLeuAsnGlnGluTrpArgValTyr 330
||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
709 GACTGTGCCTTTCTTCTGCTATTTGGGGATGTTTCAAAACCGGAAGATA 758
330 rThr.....TyrLys.....TyrPheLysLysGluGluLul 341
||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
759 TAGTGGCGACAACATAATCTACAAACCCAGGCGATCTGCTCCGGATA 808
341 ysLeuPheLeuAlaProIleSerAspSerAspGluLeuTyrAsnThrIle 357
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
809 TGGTGTACTTGGGAGCTATCACAAATTTTGCATGACTTACAATTCGATT 858
358 GlnIleLysGluTyrAspGluGlnProThrTyrSerCys..... 370
||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
859 CAAGATAAA.....TGTGTTCTCTCTGT 881
371 ....GlnLeuPheLysAspGluGluSerThrAspGluIleGlyL 386
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
882 CCAGAAATACATTTGAAATGGAGAGGAATTGCACAGAGAA...GGAC 928
386 eulLeGlyIleHis 390
||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
929 TGGCCTTTTCTCAT 942
seq_name: gb_gss:BH150480
```

```
seq_documentation_block:
LOCUS BH150480 907 bp DNA linear GSS 27-AUG-2001
DEFINITION ENT0957FB Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, DNA sequence.
ACCESSION BH150480
VERSION BH150480.1 GI:15312077
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
Eukaryota; Entamoebidae; Entamoeba.
REFERENCE 1 (bases 1 to 907)
AUTHORS Loftus.B., Wang,Z., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library (2001)
JOURNAL Unpublished (2001)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0208
Fax: 301 838 3543
Email: bjoftus@igr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
```

```
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 54
High quality sequence stop: 784.
FEATURES
source
1..907
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: PHOS1; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,
C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Barell, Oxford University Press, 1999)."
```

BASE COUNT 382 a 91 c 149 g 285 t
ORIGIN

alignment_scores:
Quality: 113.00 Length: 349
Ratio: 0.661 Gaps: 17
Percent Similarity: 48.997 Percent Identity: 19.484

alignment_block:
US-09-910-186A-8 x BH150480 ..
Align seg 1/1 to: BH150480 from: 1 to: 907

```
29 GlyTyrGlyAlaLysValGluValTyrAspGlyValGluLeuAsn..... 43
||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
26 GGTTCCTATATCGTACAACAATTCGGGATTACATTTTCATCGTTTCAT 75
44 ....AspLysAsnGlnPheLysLeuThrSerSerAlaAsnSerLys.Ile 58
||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
76 TATTGCAAAATCCTGTTTAAACTAATATAAAAGAAATTCACACTTG 125
59 ArgValThrGlnAsnGlnAsnIlePheAsnSerValPheLeuAspPh 75
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
126 AAATAGCATTAGATAAAAAAATGAGTATATCATTTACTATTATAAATAAT 175
75 eSerValSerPheTrp.....IleArgIleProLysTyrLysAsnAspG 90
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
176 TCTGTGTATATTCAGTCCTCTATAAAACCCACGTTTAGTTCATGTTT 225
90 lTleGlnAsnTyrIleHisAsnGluTyrThrIleIleAsnCysMetLys 106
||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
226 TTATTCAAAGAGGAATCATTTATGAATATGGAACATAACAACAAATGAAA 275
107 AsnAsnSerGlyTrpLysIleSerIleArgGlyAsnArgIleIleTrpTh 123
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
276 GGTGATAAT.....GGTGAAACAGAAATTCACAATA 304
123 rLeuIleAspIleAsnGlyLysThrLysSerValPhePheGluTyrAsnI 140
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
305 TTCTGTTGATATAAACCAATATGTTTCATTTGATTTCATGATCAACAA 354
140 leArg.....GluAspIle.....Ser 145
||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
355 TAGAAAGTGTGAATGCAAGAGAGATGATTTGATTGTTTCAACAAACAAT 404
146 GluTyrIleAsnArgTrp.....PhePheValThrIleThrAsnAs 159
::||| ::||| ::||| ::||| ::||| ::||| ::||| ::|||
405 AAATATATTGAAACATGGAGGATAAATTCATCTGTTTACAGGGTCTAATGA 454
```

```

159 nLeuAsnAlaLysIleTyrIleAsnGlyLysLeuGluSerAsnThrA 176
160 ::::::::::::::::::::
455 AATGTGGAAACCTTCATGTACATTAATAAGAAGA..... 488
176 spIleLysAspIleArgGluValIleAlaAsnGlyGluIleIlePheLys 192
177 ::::::::::::::::::::
489 .....ATATTGAATATTGAATTAATGATGAA 515
193 LeuAspGlyAspIleAspArgThrGlnPheIleTrpMetLysTyrPhe 209
194 ::::::::::::::::::::
516 AAAGATGGA.....TATATATATA 535
209 rIlePheAsnThrGluLeuSerGlnSerAsnIleGluArgTyrLysI 226
210 ::::::::::::::::::::
536 TTTACTTAGACAGAG.....AAAG 555
226 leGlnSerTyrSerGluTyrLeuLysAspPheTrpGlyAsnProLeuMet 242
227 ::::::::::::::::::::
556 TG...TCATTTAATGAATAT..... 572
243 TyrAsnLysGluTyrTyrMetPheAsnAlaGlyAsnLysAsnSerTyrI 259
244 ::::::::::::::::::::
573 TTTATGAGGCAAAATTAATTTATATCAAGAATATGAGGA..... 617
259 eLysLeuLysLysAspSerProValGlyGluIleLeuThrArgSerLys 276
260 ::::::::::::::::::::
618 .....T 618
276 yrAsnGlnAsnSerLysTyrIleAsnTyrArgAspLeuTyrIleGlyGlu 292
277 ::::::::::::::::::::
619 ATAATAAAGAAATAAATAAAGGATTAGAAAGATACATTTCAAGAA 668
293 .LysPheIleIleArgArgLys.SerAsnSerGlnSerIleAsnAspAsp 308
294 ::::::::::::::::::::
669 GAAAGTATATATGATGCAATCAACAGCTTAATATAGTTCATTTCAACAG 718
309 IleValArgLysGluAspTyrIleTyrLeuAspPhe..... 320
310 ::::::::::::::::::::
719 GTAGTACAAAGAGTGAATATATCGAAGTGAAACACACAGAAAAACATT 768
321 .....PheAsnLeu.....AsnGlnG 326
322 ::::::::::::::::::::
769 GCATTCATATATTAAGAACACCAATTTGGGTGTATAATGGTGACATGATG 818
326 luTrpArgValTyrThrTyrLysTyrPheLysLysGluGlu 339
819 ATATTTATGTATAGTATAGTATCTTATTAAGGAAGTGAG 859

```

seq_name: gb_est2:BM276398

seq_documentation_block:

CDS BM276398 439 bp mRNA linear EST 20-DEC-2001
DEFINITION PFESToa82f02.y1 Plasmodium falciparum 3D7 gametocyte cDNA library
Plasmodium falciparum 3D7 cDNA 5', mRNA sequence.

ACCESSION

VERSION BM276398

KEYWORDS

SOURCE

ORGANISM

Plasmodium falciparum 3D7.

Plasmodium falciparum 3D7.

Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.

1 (bases 1 to 439)

Tang,K., Cole,R., Chakrabarti,D., Haywood,R., Clifton,S., Pape,D.,

Marra,M., Hillier,L., Martin,J., Wyllie,T., Dante,M., Theising,B.,

Bowers,Y., Gibbons,M., Ritter,E., Bennett,J., Jentes,E., Ronko,I.,

Tsagareishvili,R., Belaygorod,L., Franklin,C., Carr,L., Grow,A.,

Maguire,L., Richey,J., Wadkins,J., Kennedy,S., Levinso,D.,

Waterston,R., Wilson,R. and Sibley,D.

WashU Plasmodium EST Project

Unpublished (2001)

Contact: L. David Sibley

WashU Plasmodium EST Project

Washington University School of Medicine

4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA

Tel: 314 286 1800
Fax: 314 286 1810
Email: est@wuston.wustl.edu
Library was constructed by R. Haywood. DNA sequencing by:
Washington University Genome Sequencing Center For information on
obtaining a clone please contact: L. David Sibley
(sibley@borcim.wustl.edu), Washington University
Seq primer: -400P from Gibco
High quality sequence stop: 430.

FEATURES:

source
1..439
/organism="Plasmodium falciparum 3D7"
/db_xref="taxon:36329"
/clone_lib="Plasmodium falciparum 3D7 gametocyte cDNA
library"
/dev_stage="gametocyte (stage III-V)"
/lab_host="DH10B (GeneHog, Invitrogen, Inc.)"
/notes="Vector: pBluescript SK plus; Site_1: EcoRI; Site_2:
XhoI; The library was constructed by R Haywood. cDNAs were
synthesized from gametocyte poly(A)+ RNA by oligo d(T)
priming, size-selected and directionally cloned into the
EcoRI (5' end) to XhoI (3' end) sites of the Uni-ZAP XR
lambda vector (Stratagene). The primary library was mass
excised as phagemid using the ExAssist helper phage
(Stratagene). Clones were mass excised using the ExAssist
helper phage (Stratagene), the phagemids were precipitated
with PEG 8000 and extracted with phenol/chloroform.
Phagemid DNA was electroporated into DH10B cells. Clone
Availability: David Sibley, Washington University."

BASE COUNT 220 a 37 c 41 g 141 t
ORIGIN

alignment_scores:

Quality: 112.50 Length: 153

Ratio: 1.424 Gaps: 5

Percent Similarity: 51.634 Percent Identity: 25.490

alignment_block:

US-09-910-186A-8 x BM276398

Align seg 1/1 to: BM276398 from: 1 to: 439

```

161 AsnAsnAlaLysIleTyrIleAsnGlyLysLeuGluSerAsnThrAspI 177
162 ::::::::::::::::::::
12 AATAATACAGTGTACATAAATGTTAAGAACTCATTTAATAATGAGAG 61
163 ::::::::::::::::::::
177 eLysAspIleArgGluValIleAlaAsnGlyGluIleIlePhe..... 191
164 ::::::::::::::::::::
62 AAATGATATAAACAATGCAAAAAAATAGTTATATCAACTATTATAATA 111
165 ::::::::::::::::::::
192 .....LysLeuAspGlyAspIleAspArgThrGlnPheIleTrpMetLys 206
166 ::::::::::::::::::::
112 AATTAATAAATAAATGGAAT.....AAG 134
167 ::::::::::::::::::::
207 TyrPhe.....SerIlePheAsnThrGluLeuSerGlnSe 218
168 ::::::::::::::::::::
135 TATATGATACCTGTCATAAATAATATATGTAAGGAACAACAGATTGA 184
169 ::::::::::::::::::::
218 rAsnIleGluArgTyrLysIleGlnSerTyrSerGluTyrLeuLysA 235
170 ::::::::::::::::::::
185 TAATATATATGAACATATATGAGTACGCTCTCCAAAAAATAATATGA 234
171 ::::::::::::::::::::
235 spPhe..... 236
172 ::::::::::::::::::::
235 ATATGAATTCATTTGGTAAAGTATACCATTAACAATATTATAAAGAATA 284
173 ::::::::::::::::::::
237 .....TrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPh 250
174 ::::::::::::::::::::
285 ATTTATGTCACAAATAATAATAAATAATATAATATTCAACAATTTCAAT 334
175 ::::::::::::::::::::
250 eAsnAlaGlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerPro 267
176 ::::::::::::::::::::

```

```

335  AAATGTGAATGAAAAAACTCTTTTCATAAATTAATAAAAGT..... 377
267  alGlyGluLeuThrSerLysTyrAsnGlnAsnSerLysTyrTle 283
378  .GGTCATATAAAACATATTCCTTTAATTAATTATAAAATCTTATGCC 425
284  AsnTyrArg 286
      ::::
426  ACATATAAA 434

```

seq_name: gb_est1:AL557255

seq_documentation.

LOCUS
DEFINITION

DEFINITION RESCUE, MRN

CESSION AL557

AL557

KEYWORDS EST.
SOURCE human

SOURCE	ORGANISM	HOMO	HUMAN
1	1	1	1
2	2	2	2
3	3	3	3
4	4	4	4
5	5	5	5
6	6	6	6
7	7	7	7
8	8	8	8
9	9	9	9
10	10	10	10
11	11	11	11
12	12	12	12
13	13	13	13
14	14	14	14
15	15	15	15
16	16	16	16
17	17	17	17
18	18	18	18
19	19	19	19
20	20	20	20
21	21	21	21
22	22	22	22
23	23	23	23
24	24	24	24
25	25	25	25
26	26	26	26
27	27	27	27
28	28	28	28
29	29	29	29
30	30	30	30
31	31	31	31
32	32	32	32
33	33	33	33
34	34	34	34
35	35	35	35
36	36	36	36
37	37	37	37
38	38	38	38
39	39	39	39
40	40	40	40
41	41	41	41
42	42	42	42
43	43	43	43
44	44	44	44
45	45	45	45
46	46	46	46
47	47	47	47
48	48	48	48
49	49	49	49
50	50	50	50
51	51	51	51
52	52	52	52
53	53	53	53
54	54	54	54
55	55	55	55
56	56	56	56
57	57	57	57
58	58	58	58
59	59	59	59
60	60	60	60
61	61	61	61
62	62	62	62
63	63	63	63
64	64	64	64
65	65	65	65
66	66	66	66
67	67	67	67
68	68	68	68
69	69	69	69
70	70	70	70
71	71	71	71
72	72	72	72
73	73	73	73
74	74	74	74
75	75	75	75
76	76	76	76
77	77	77	77
78	78	78	78
79	79	79	79
80	80	80	80
81	81	81	81
82	82	82	82
83	83	83	83
84	84	84	84
85	85	85	85
86	86	86	86
87	87	87	87
88	88	88	88
89	89	89	89
90	90	90	90
91	91	91	91
92	92	92	92
93	93	93	93
94	94	94	94
95	95	95	95
96	96	96	96
97	97	97	97
98	98	98	98
99	99	99	99
100	100	100	100

Eukaryotes

Маппа

REFERENCE

AUTHORS
TITLE
LI, W.
Full-

JOURNAL
Full. Unpub.

COMMENT

190IlePheLysLeuAspGlyAspIleAspAr 199
 314 TTCAAGTCAGATGTTGCATCCAAATTTTGGGAAGCGTCCGATGCATTAA 363
 199 gThrGlnPheIleTrpMetLysTyrPheSerIlePheAsnThrGluLeu 216
 364 GGAAGAAATTTCCAAATGAAATCAAGTAGTGTGTCCAGAGTTGATGTG 413
 216 erGln...SerAsnIleGluGluGlyTyrLysIleGlnSerTyrSerGlu 231
 414 ATCAGCACTCTGCATAGTCCAGAGATACAGGATAAGCAAAATACCAACC 463
 232 TyrLeuLysAspPheTrpGlyAsnProLeuMetTyrAsnLysGluTyrTy 248
 464 ...CTCAAAATGTTTT...CGTAATGGGATGATGATGAAGAGAGAATC... 505
 248 rMetPheAsnAlaGlyAsnLysAsn.....SerTyrIleL 260
 506AGGGGTGAGCGATCAGTCAAAAGCATTTGCCAGATTACATCA 545
 260 ysLeuLysLysAspSerProValGlyGluIle..... 270
 546 GGCAACAAAAAGTGCACCCATTCAGAGAAATTCGGGACTTAGCAAAATC 595
 271LeuThrArgSerLys.....TyrAsnGlnAsnSe 280
 596 ACCACTCTTGATCGACGCAAAAGAAATATCATTTGGATATTTTGACAAA 645
 280 rLysTyrIleAsnTyrArgAspLeuTyrIleGlyGluLysPheIleLea 297
 646 GGACTCGGACAACTATAGA.....GTTTGTG 671
 297 rgArgLysSerAsnSerGlnSerIleAsnAspIleValArgLysGlu 313
 672 AACGAGTAGCGAAT.....ATTGTCATGAT 697
 314 AspTyrIleTyrLeuAspPhePheAsnLeuAsnGlnGluTyrArgValTy 330
 698 GACGTGCGCTTCCTTCGCAATTTGGGAGATGTTTCMAAACCGGAAGATA 747
 330 rThr.....TyrLys...TyrPheLysLysGluGluL 341
 748 TAGTGGGCGAACATAATCTCAAAACCAACGAGGCAATTCGTCCGGATA 797
 341 ysLeuPheLeuAlaProIleSerAspSerAspGluLeuTyrAsnThrIle 357
 798 TGGTGTACTTGGGAGCATGACAAATTTGATGTGACTTCACAATGGGATT 847
 358 GlnTleLysGluTyrAspGluGlnProThrTyrSerCys..... 370
 848 CAAGATAAA.....TGTTCTCTCTGT 870
 371GlnLeuLeuPheLysLysAspGluGluSerThrAspGlu 383
 871 CCGAGAAATACATTTGAAATGGAGAGCAATTCAGACAAGAA 913

seq_name: qb qss:AZ524527

seq documentation block:

LOCUS A2524527

DEFINITION 233pbC08 Pb MBN #21 Plasmodium berghei genomic 3', DNA sequence.

ACCESSION AZ524527

```

VERSION
KEYWORDS
CSS
AZ524527.1
GI:13964499

```

KEYWORDS
SOURCE
GSS:
place

ORGANISM
SOURCE
plasmodium berghei
FASMODIAN berghei
FASMODIAN berghei

Eukaryota; Alveolata

REFERENCE

AUTHORS Carlton, J.M.-R. and

TITLE The Plasmodium viva

JOURNAL
Parasitol. Today 16

COMMENT Contact: Dame JB

Dept. of Pathobiology, College of Veterinary Medicine
University of Florida
2015 SW 23rd Avenue, Bldg 1017, Gainesville, FL 32611, USA
Tel: 352 392 4700
Fax: 352 392 9704
Email: damej@mail.vetmed.ufl.edu
Seq primer: M13(-20) forward
Class: shotgun.

```

FEATURES             source
1..642
/organism="Plasmodium berghei"
/strain="ANKA clone 15cyl (clone of the ANKA 8417 clone)"
/db_xref="taxon:5821"
/clone_lib="Pb MBN #21"
/dev_stage="asexual blood forms"
/lab_host="Mus musculus"
/note="Vector: pBluescript SK(+) vector DNA, phagemid
excised from lambda ZAP; Site_1: EcoRV; Site_2: EcoRV;
Genomic DNA was prepared from asynchronous blood stage
forms of the cloned ANKA isolate of P. berghei grown in
laboratory Swiss white mice. The DNA was purified from
contaminating host DNA by Hoechst Dye 33258-CsCl
ultracentrifugation and precipitated. Purified DNA was
digested with mung bean nuclease in the presence of 36-38%
formamide at 50 C, as described (Vernick, K.D., Imberski,
R.B., and McCutchan, T.F. 1988. Nucleic Acids Research
16:6883-6896). The ends of the digestion fragments were
polished using T4 DNA polymerase, and the fragments size
selected in the range 500-2000 bp. These were ligated into
the EcoRV-cleaved and dephosphorylated pBluescript SK(+)
vector. Recombinant plasmids were used to transform E.
coli XL10-Gold host cells."
```

BASE COUNT 222 a 85 c 76 g 258 t 1 others
ORIGIN

```

alignment_scores:
  Quality: 112.00      Length: 227
  Ratio: 1.067        Gaps: 12
Percent Similarity: 46.256 Percent Identity: 24.229

alignment_block:
US-09-910-186a-8 x AZ524527/rev ..

Align seg 1/1 to reverse of: AZ524527 from: 1 to: 642

38 AspGlyValGluLeuAsnAspLysAsnGlnPhelys.LeuThrSerSera 54
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
631 GATCATAAAGAGATAAATCGTACTAACCAATCGAAGGTGCTACTCTG 582

54 laAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePheAsnSer 70
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
581 GTAAGCT.....CAACACATAACCAACTTGCT 553

71 ValPhe.....LeuAsp.....PheSerValSerPheTrpIleAr 82
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
552 ATAACATCAAGATTAGATTACCAAAATTATAGACATGGTGTACTTTCAG 503

82 gIleProLysTrpLysAsnAspGlyIleGlnAsnTrpIleHis....AsnG 98
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
502 ATTGCTTAACATATAATGAAGCAATGAAGAATGTTAAGAATGATAATA 453

98 luTyThrIleIleAsnCysMetLysAsn..... 107
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
452 ATGTTTCAATAGTACATAAAATTAATTTTCAAGAAATGATGGCCCTATA 403

108 ...AsnSerGlyTrpLysIleSerIleArgGlyAsnArgIleIleTrpTh 123
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
402 TCAATATAAGCTACAAATCAGATAGGC.....GTGTATGC 365

123 rLeuIle.....AspIleAsnGlyLysT 131
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
364 CTTATTANATTCAATGAAAAAATATTTTAAATGATATTATAATAACT 315
```

```

131 hrLysSerValPhePheGluTrpAsnIleArgGluAspIleSer..... 145
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
314 TAAACACAGAAATTTATAAATTTTAAACACAGTTGATGATTATCGTTTAAAC 265

146 GluTrpIleAsnArgTrp.....PhePheValThrIleThrAs 158
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
264 ATATATTTTAAAGAATGTTATTAATAATAGTTTATTTTATAGCTCTTGACA 215

158 nasLeuAsn..... 161
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
214 CCATTGAATTTATTCCTTTTGAATAATAATATGTTTAAAAATAATAAAA 165

162 .....AsnAlaLysIleTrpIle 167
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
164 ATGAAGAATTTAGAAATATATTATGGCAAAATTAACAATAAAATATTGTA 115

168 AsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGluValI 184
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
114 TATTACATGCAGAGGCTAATGATGATGATAAATATGTCACAACCGTTTC 65

184 eAlaAsnGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThrG 201
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
64 GCTCGAAATTTTATTCCTTTTATAAATGCCT..... 34

201 lnPheIleTrpMetLysTrpPheSerIle 210
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
33 .....ATGAAATTTTAAAGATT 16

seq_name: gb_est2:BG603391
seq_documentation_block:
LOCUS BG603391 603 bp mRNA linear EST 14-AUG-2001
DEFINITION EST502481 Plasmodium yoelii sporozoite cDNA Plasmodium yoelii CDNA
clone PYCDRL3, mRNA sequence.
ACCESSION BG603391 GI:15153405
VERSION BG603391
KEYWORDS EST.
SOURCE Plasmodium yoelii.
ORGANISM Plasmodium yoelii.
REFERENCE 1 (bases 1 to 603)
AUTHORS Kappe,S.H.I., Gardner,M.J., Brown,S.M., Ross,J., Matuschewski,K.,
Ribeiro,J.M., Adams,J.H., Quackenbush,J., Cho,J., Carucci,D.J.,
Hoffman,S.L. and Nussenzweig,V.
Exploring the transcriptome of the malaria sporozoite stage
Proc. Natl. Acad. Sci. U. S. A. 98 (17), 9895-9900 (2001)
2136555
Contact: Malcolm J. Gardner
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Drive, Rockville, MD 20850, USA
Tel: 301 838 3519
Fax: 301 838 0208
Email: gardner@tigr.org
Request for clones, please contact: Stefan Kappe,
kappes01@popmail.med.nyu.edu Michael Heidelberg Division,
Department of Pathology New York University School of Medicine.
FEATURES
source
1..603
/organism="Plasmodium yoelii"
/strain="17XNL"
/db_xref="taxon:5861"
/clone="PYCDRL3"
/clone_lib="Plasmodium yoelii sporozoite cDNA"
/dev_stage="sporozoites from salivary gland"
/lab_host="E. coli TOP10"
/note="Vector: pCR4; TA cloning; Plasmodium yoelii
sporozoite cDNA library from salivary gland sporozoites 14
days post-infection"
BASE COUNT 337 a 30 c 46 g 190 t
ORIGIN
```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2002, 15:01:23 ; Search time 1635.05 Seconds
(without alignments)
11069.641 Million cell updates/sec

Title: US-09-910-186A-7

Perfect score: 1341

Sequence: 1 gaattcacgatgcccaacaa.....ggaccgaatagtagaattc 1341

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:**
2: em_esthum:**
3: em_estin:**
4: em_estmu:**
5: em_estov:**
6: em_estpl:**
7: em_estro:**
8: em_htc:**
9: gb_est1:**
10: gb_est2:**
11: gb_htc:**
12: gb_gss:**
13: em_gss_hum:**
14: em_gss_inv:**
15: em_gss_pln:**
16: em_gss_vrt:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	70.6	5.3	827	12	CNS02156
2	59.4	4.4	394	9	AU060224 Tetraodon
3	59.2	4.4	641	10	AU060224 AU060224
4	57.4	4.3	849	12	BM181884 f551b11.y
5	54.6	4.1	542	10	AZ546009 ENTFW53TF
6	53.2	4.0	575	10	BI594734 As_tgz_43
7	52.4	3.9	500	12	BM278279 As_tgz_55
8	52.2	3.9	612	9	B67199 CpG0015B Cp
9	52	3.9	970	12	BE195101 HVSMEH08
10	51.8	3.9	681	12	AL243904 Tetraodon
11	51.8	3.9	904	12	AL193990 Tetraodon
12	51.6	3.8	546	10	BM278263 As_tgz_55
13	50.8	3.8	519	10	BM278164 As_tgz_54
14	49.6	3.7	615	9	AU061402 AU061402
15	49.2	3.7	522	10	BM278558 As_tgz_64
16	48.8	3.6	511	10	BM004465 TgESTzya8
17	48.8	3.6	905	12	AZ550256 ENTEV58TR

18	48.6	3.6	450	9	AU060996
19	48.6	3.6	795	12	AZ528485
20	48.6	3.6	823	12	AZ676218
21	48.6	3.6	843	12	BM139532
22	48.2	3.6	900	12	AZ549980
23	48	3.6	628	9	AU060230
24	48	3.6	878	12	BM153470
25	47.8	3.6	1016	12	CNS04ESQ
26	47.6	3.5	445	10	BM167469
27	47.6	3.5	598	10	BM170666
28	47.6	3.5	605	10	BM163520
29	47.6	3.5	645	10	BM163530
30	47.6	3.5	647	10	BM163120
31	47.6	3.5	671	10	BM160252
32	47.6	3.5	694	10	BM168242
33	47.6	3.5	717	10	BM160500
34	47.6	3.5	747	10	BM162732
35	47.6	3.5	756	10	BM162492
36	47.6	3.5	794	10	BM159906
37	47.2	3.5	477	10	BM278697
38	47.2	3.5	529	10	BM278587
39	46.8	3.5	546	10	C22974
40	46.6	3.5	645	9	AI389106
41	46.6	3.5	805	12	AZ753057
42	46.4	3.5	525	10	BM278174
43	46.4	3.5	890	12	AZ530768
44	46.2	3.4	1101	12	CNS0039G
45	46	3.4	548	10	BM159818

ALIGNMENTS

RESULT 1

CNS02156

LOCUS

DEFINITION

CNS02156 827 bp DNA linear GSS 12-MAY-2000
Tetraodon nigroviridis genome survey sequence PUC-ori end of clone
224F10 of library G from Tetraodon nigroviridis, genomic survey
sequence.

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

1. 827

/organism="Tetraodon nigroviridis"

location/Qualifiers

http://www.genoscope.cns.fr/Tetraodon.

genome. For more information, please take a look at

scale clone-end sequencing project of the Tetraodon nigroviridis

This sequence is a single read and was generated as part of a large

Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases

Genoscope.

3 (bases 1 to 827)

Unpublished

Tetraodon nigroviridis DNA sequence

Human gene number estimate provided by genome wide analysis using

Saurin, W. and Weissbach, J.

Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,

Roest-Crollius, H., Jaillon, O., Dasilva, C., Bouneau, L., Fisher, C.,

1 (bases 1 to 827)

1

```
/db_xref="taxon:99883"
/clone="224F10"
/clone_lib="C"
/notes="Genoscope sequence ID : COAG224DC05SP1-end :
PUC-ori"
BASE COUNT      368 a   203 c   106 g   121 t   29 others
ORIGIN

Query Match      5.3%; Score 70.6; DB 12; Length 827;
Best Local Similarity 46.5%; Pred. No. 9.5e-08;
Matches 179; Conservative 13; Mismatches 193; Indels 0; Gaps 0;

QY 266 acaagaacgagcggtatccagaattacatccaaatgaatacaccatcataactgcgatga 325
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 186 ACAACACAAATAACAAATACAGCAATACACAAATAACACAAATAACACAAATAACACAA 245

QY 326 agaataactctgttggaagatctccatccgcggtgaaccgtatctatctggactctgatcg 385
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 246 ACAATACAGCATAACAAATAACAAATAACAAATAACAAATAACAAATAACAAATAACAGCS 305

QY 386 atatcaacggtgaagacaaatctgtattcttctgaatacaacatccggtgaagacatctctg 445
      ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 306 ATAACAACAATAACAGCATAACAAATAACAAATAACAAATAACAAATAACAAATAACAA 365

QY 446 aatacataatcgctgttcttctgttaccatcaccataaaccgtgaacaatgctaaatct 505
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 366 ACACSTATACAAATAACAGCVATACACAAATAACACAAATAACACAAATAACAGTGATACAA 425

QY 506 acatcaacggtgaactggaatcttaataccgacatcaaaagacatccgtgaagtctatcgcta 565
      ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 426 ATAACAACAATAACAAATAACAGCAGATACACAAATAACAGCAGATACACAAATAACAGC 485

QY 566 acggtgaatcttcaactggagcggtgacatcgatctgacccagttctatctctggatga 625
      ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 486 ACAACAACAACAACAATAACAGCGATAACAGCATAACAGCATAACAAATAACAGTGTATAACA 545

QY 626 aatactctccatcttcaaccgga 650
      ||| ||| |||| |||| |||
Db 546 ACAATACAGCGATGACAAACAATRA 570

RESULT 2
AU060224
LOCUS
DEFINITION
Dictyostelium discoideum SL (H.Urushihara) Dictyostelium
discoideum cDNA clone SLA610, mRNA sequence.
ACCESSION
AU060224
VERSION
AU060224.1 GI:4881328
KEYWORDS
EST.
SOURCE
Dictyostelium discoideum.
ORGANISM
Dictyostelium discoideum
Eukaryota; Mycetozoa; Dictyostellida; Dictyostellum.
REFERENCE
1 (bases 1 to 394)
AUTHORS
Morio,T., Urushihara,H., Saito,T., Ugawa,Y., Mizuno,H., Yoshida,M.,
Yoshino,R., Mitra,B.N., Pi,M., Sato,T., Takemoto,K., Yasukawa,H.,
Williams,J., Maeda,M., Takeuchi,I., Ochiai,H. and Tanaka,Y.
Developmental cDNA in Dictyostelium discoideum
Unpublished (1998)
TITLE
Contact: Hideko Urushihara
Institute of Biological Sciences
University of Tsukuba
3-3-10 Ten-nodai, Tsukuba, Ibaraki 305, Japan
Email: d402hu@akura.cc.tsukuba.ac.jp
PROJECT - Dictyostelium discoideum cDNA project in Japan.
FEATURES
source
1..394
/organism="Dictyostelium discoideum"
/strain="AX4"
/db_xref="taxon:44689"
/clone="SLA610"
/clone_lib="Dictyostelium discoideum SL (H.Urushihara)"
/dev_stage="slug"
```

```
BASE COUNT      223 a   134 c   8 g   28 t   1 others
ORIGIN

Query Match      4.4%; Score 59.4; DB 9; Length 394;
Best Local Similarity 52.2%; Pred. No. 6.4e-05;
Matches 132; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 266 acaagaacgagcggtatccagaattacatccaaatgaatacaccatcataactgcgatga 325
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 131 ACACACCAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 190

QY 326 agaataactctgttggaagatctccatccgcggtgaaccgtatctatctggactctgatcg 385
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 191 ACAACACAAACAACAACAACAACAACAACAACAACAACAACAACAATATCATCAATCAGCAACA 250

QY 386 atatcaacggtgaagacaaatctgtattcttctgaatacaacatccggtgaagacatctctg 445
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 251 ATATCATCAATATACAAACAATACCATCAACCATCAACAACAACAACAACAACAACAACA 310

QY 446 aatacataatcgctgttcttctgttaccatcaccataaaccgtgaacaatgctaaatct 505
      ||||| |||| |||| |||| |||| |||| |||| |||| |||| |||| |||| ||||
Db 311 ACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 370

QY 506 acatcaacggtga 518
      ||||| |||| |||| |||
Db 371 ACAACAACAACA 383

RESULT 3
BM181884
LOCUS
DEFINITION
fm5lb1l.y1 Sugano SJD adult male Danio rerio cDNA clone 5412044 5'
similar to contains element TAR1 repetitive element ;, mRNA
sequence.
ACCESSION
BM181884
VERSION
BM181884.1 GI:17512842
KEYWORDS
EST.
SOURCE
zebrafish.
ORGANISM
Danio rerio
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes
; Cyprinidae; Danio.
REFERENCE
1 (bases 1 to 641)
AUTHORS
Clark,M., Johnson,S.L., Lehrach,H., Lee,R., Li,F., Marra,M., Eddy
,S., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T., Underwood
,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B.,
Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E.,
Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R.
and Wilson,R.
WashU Zebrafish EST Project 1998
Unpublished (1998)
TITLE
Contact: Stephen L. Johnson
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: zbrafish@wustl.edu
Library constructed by Dr. Sumio Sugano and Dr. Koichi Kawakami DNA
Sequencing by: Washington University Genome Sequencing Center clone
distribution information can be found through the I.M.A.G.E.
Consortium/LLNL, send email to: info@image.llnl.gov
Seq primer: T3 ET from Amersham
High quality sequence stop: 527.
FEATURES
source
1..641
/organism="Danio rerio"
/db_xref="taxon:7955"
/clone="5412044"
/clone_lib="Sugano SJD adult male"
/sex="male"
/tissue_type="whole body"
/dev_stage="adult"
```


Library was made from dissected testis germinal zone from adult male *Ascaris suum* collected from abattoirs. Constructed by Michelle Lizotte-Waniewski for Alan Scott, Johns Hopkins University Medical School, Baltimore, MD.

```
BASE COUNT      204 a 148 c 74 g 149 t
ORIGIN

Query Match      4.0%; Score 53.2; DB 10; Length 575;
Best Local Similarity 48.1%; Pred. No. 0.0032;
Matches 182; Conservative 0; Mismatches 193; Indels 3; Gaps 1;

QY 273 cgagcggtatccagaattacatccacatgaataacacacatcatcaactgcgatgaagataa 332
Db 54 CGTCGGGTCCTCAAAAGACCCACTGGAATAGCAACAGCAACTTCAATTGTTTTTCAACAA 113
QY 333 ctctggttggaagatctccatccgcgggtaaccgtatcatctgactctgctgatatacaa 392
Db 114 CAACAATTTCAACAACAGACCCCTCAACAACAACACTTCAACAGCAATTTCAACAACAGCAA 173
QY 393 cggtaagaccacaatctgtattcttcgaatacaacatcccgtaagacatctctgaatacat 452
Db 174 CTTCACAGCAACTTCATTTTATCTTTTATGTCCTCAACAACTTTTGTGGAGAG 233
QY 453 caatcgctggttcttctgtaccatccacaaataacactgaacatgtctaaatctacatcaa 512
Db 234 CACCTTCAGTAGCAGC---AGCTTCAGCAATTACCTCAACAACATATTTTAACAACCTCAA 290
QY 513 cggtaactggaatcttaataaccgacatcaagacatcccgtaaggttatcgctgaacggtga 572
Db 291 CTATAACTTGAACAATCTCAACAACCTCTAATAACAACCTCAACAATCTCAACAACAACCT 350
QY 573 aatcatcttcaactgagcggtgacatcgatctacccagttcatctctggtgaatactt 632
Db 351 CAGTAACCTCAACAACACCTTAGCAACCTCAGCAGACAGTAAGTAACCTCAACAACAGCAACT 410
QY 633 ctccattcttcaacacaga 650
Db 411 CAACAGCAACTTCAACAA 428

RESULT 7
LOCUS      B67199
DEFINITION Cpg0015B CpIOWAGDNa1 Cryptosporidium parvum genomic, DNA sequence.
ACCESSION  B67199
VERSION     B67199.1 GI:2642691
KEYWORDS   GSS.
SOURCE     Cryptosporidium parvum.
ORGANISM   Cryptosporidium parvum
            Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida;
            Cryptosporidiidae; Cryptosporidium.
REFERENCE  1 (bases 1 to 500)
AUTHORS   Strong, W.B. and Nelson, R.G.
TITLE     Preliminary profile of the Cryptosporidium parvum genome: an
            expressed sequence tag and genome survey sequence analysis
JOURNAL   Mol. Biochem. Parasitol. 107 (1), 1-32 (2000)
MEDLINE   20183851
COMMENT   Contact: Nelson, R. G.
            Depts. of Medicine & Pharmaceutical Chemistry
            San Francisco General Hospital-University of California, San
            Francisco
            Box 0811, San Francisco, CA 94143-0811, USA
            Tel: 415 206 8846
            Fax: 415 206 3353
            Email: malaria@itsa.ucsf.edu
            Submitted sequence has been edited to remove vector sequences 5' to
            the insert, to correct miscalled bases and assign uncalled (N)
            bases throughout the sequence, and to terminate when base-calling
            became ambiguous.
            Seq primer: T3
            Class: shotgun
            High quality sequence stop: 500.
```

FEATURES

source

Location/Qualifiers
1..500
/organism="Cryptosporidium parvum"
/strain="IOWA"
/db_xref="taxon:5807"
/clone_lib="CpIOWAGDNa1"
/lab_host="E. coli XL2 Blue MRF"
/note="Vector: pBlueScript II (SK-); Site 1: EcorV; C.
parvum (IOWA isolate) genomic DNA was hydrodynamically
sheared to produce fragments having a tight size
distribution between 2-4 kb by Dr. Ivonne Thorstenson of
the Stanford DNA Sequencing and Technology Center
(http://sequence-www.stanford.edu/group/techdev/shear.htm
). The randomly sheared gDNA was chromatographed on
Sephacryl S-400 to remove any small fragments and DNA
eluting in the void volume was subcloned into an Ecor
V-digested, alkaline phosphatase-treated pBlueScript II
(SK-) vector and transformed into E. coli strain XL2 Blue
MRF'. Recombinant clones from the first plating of the
library were selected for sequence analysis using T3 and
T7 primers."

BASE COUNT 221 a 195 c 12 g 72 t
ORIGIN

Query Match 3.9%; Score 52.4; DB 12; Length 500;
Best Local Similarity 45.4%; Pred. No. 0.0049;
Matches 188; Conservative 0; Mismatches 226; Indels 0; Gaps 0;

QY 625 aaatactctccatcttcaaacaccgaactgtctcagtcctcaatcgaagaacggtacaag 684
Db 83 AACTCAACAACACTACAACAACACTACCACTACGACTACCTACTTACTACAACAACACTACTAC 142
QY 685 atccagttcttactccgaataactcctgaaagactctctggtggttaactccgctggtgatacaacaaa 744
Db 143 TACCACCACTACTACGACACGACACCACTACCACTACTACAACAACACTACTACTACTAAC 202
QY 745 gaatactatgttcaatgctggtgaagaactcttacctcaactgaagaagactct 804
Db 203 CACCCTACTACAACAACAACACTACTACCCTACTACTACAACAACAACACTACTACTACTAC 262
QY 805 ccggttggtgaatcctcgtactcgttccaaatacaaacacagaaactcttaatacatcaactac 864
Db 263 TACTACTACGACACAACAACACTACTACCCTACTACCACAACAACAACAACAACAACAACACTACTAC 322
QY 865 cgcgaactgtacatcgggtgaaaagttcactccgctcgcaaatcttaactctcgtccatc 924
Db 323 TACCCTACTACTACTACACCACTACCACTACTACGACACCAACAACAACAACAACAACAACAAC 382
QY 925 aatgatgacatcgtacgtaaagaagactacatctactcgtgactcttcaacctgaatcag 984
Db 383 TACTACAACCAACAACCACTACTACGACTACTACAACCACTACTACAACCACTACTACTACTAC 442
QY 985 gaatgctggtatcacacctcaagaagtcttcaagaagaagaagaagcttttc 1038
Db 443 TACTACCACAACAACAACAACACTACTACTACTACTACGACGACGACTACCCTACTAC 496

RESULT 8
LOCUS BE195101
DEFINITION HVSMEH0088E19f Hordeum vulgare 5-45 DAP spike EST library
HVCNDA0009 (5 to 45 DAP) Hordeum vulgare cDNA clone HVSMEH0088E19f,
mRNA sequence.
ACCESSION BE195101
VERSION BE195101.3 GI:16321082
KEYWORDS EST.
SOURCE barley.
ORGANISM Hordeum vulgare
 Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooideae
 ; Triticeae; Hordeum.
REFERENCE 1 (bases 1 to 612)


```
Db 598 AGCAGCTACATCAACAAACAGCAGCBACAAACAAACAGCAGCAGCAACAAACAGCAAC 657
QY 502 atctacatacaaggttaactggaatctaataccgacatcaaaagacatcgc 551
Db 658 AACAACAGCAGCAGCGGACAAACAAACAAACAGCAGCAGCAACAAACAGCAG 707

RESULT 10
LOCUS CNS02EOD 681 bp DNA linear GSS 13-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence T7 end of clone
262H14 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL193990
VERSION AL193990.1 GI:7832096
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 681)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 681)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Barnot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 681)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
source
1..681
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="262H14"
/clone_lib="G"
/note="Genoscope sequence ID : C0AG262DD07Lp1-end : T7"
BASE COUNT 74 a 105 c 170 g 282 t 50 others
ORIGIN

Query Match 3.9%; Score 51.8; DB 12; Length 681;
Best Local Similarity 42.2%; Pred. No. 0.0081;
Matches 164; Conservative 17; Mismatches 208; Indels 0; Gaps 0;

QY 262 aaatacaagaacgaggtatccagaattaccatcccaatgaatacaccatcaactgc 321
Db 399 AACWACAACAAAAACACACACACACACACACACACACACACACACACACACACAC 340
QY 322 atgaagaataactctggttggaagtctccatcccggttaaccggtatctatcgtgactctg 381
Db 339 GACACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 280
QY 382 atcgatatacaaggttaagaccacaaatctgtattcttcgaatacaacatccggtgaagacatc 441
Db 279 GACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 220
QY 442 tctgaatacatcaatcgctggttcttcgtttaccatcaccaataaacctgaacaaatgctaaa 501
```

```
Db 219 AAAAAAACAACAACRCGCAACNACGCGNACAACAACNACAACAACAAAAACRACAAC 160
QY 502 atctacatacaaggttaactggaatctaataccgacatcaaaagacatccggtgaagtcttc 561
Db 159 AACRACRCGCGACAAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 100
QY 562 gctaacggtgaaatcatcttcaactggacggtgacatcgatcgaccagttcattctg 621
Db 99 RACAACRACRCGCGCAACAACAACAACAACAACAACAACAACAACAACAACAACAACA 40
QY 622 atgaatacttccatcttcaaccaccga 650
Db 39 RACAACRACAGCNAACAGCAACRACAAC 11

RESULT 11
LOCUS CNS03HAO 904 bp DNA linear GSS 17-MAY-2000
DEFINITION Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
026K02 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION AL244017
VERSION AL244017.1 GI:7965029
KEYWORDS GSS; genome survey sequence.
SOURCE Tetraodon nigroviridis.
ORGANISM Tetraodon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Acanthomorpha; Acanthopterygii; Percomorpha; Tetraodontiformes;
Tetraodontidae; Tetraodon.
REFERENCE 1 (bases 1 to 904)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizames,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 904)
AUTHORS Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Barnot,A., Fizames,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence
JOURNAL Unpublished
REFERENCE 3 (bases 1 to 904)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.
FEATURES
source
1..904
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="026K02"
/clone_lib="G"
/note="Genoscope sequence ID : C0BC026BF01SP1-end :
PUC-Ori"
BASE COUNT 302 a 276 c 177 g 133 t 16 others
ORIGIN
```

```
Query Match 3.9%; Score 51.8; DB 12; Length 904;
Best Local Similarity 48.7%; Pred. No. 0.009;
Matches 167; Conservative 1; Mismatches 173; Indels 2; Gaps 1;

QY 262 aaatacaagaacgaggtatccagaattaccatcccaatgaatacaccatcaactgc 321
Db 298 AAATTCATCACCATTGACAAACAGCAACAAACAAACAGAGTGACAAACAGCAACAGCAGC 357
```

```

Qy 322 atgaagaataactctggttggaagatctccatccgcggtgaacggtatcatctggaactctg 381
Db 358 MACAACAGTAGTAATAAGCAACCAACAGCAACCAACCAACCAACCAACCAACCAACCAAC 417
Qy 382 atcgatataacggtgaagacaaatctgta--ttcttcgaatacaacatccggtgaagaca 439
Db 418 AACAAACAAACAGCAACCAACAAACAGCAACCAACAAACAAACAAACAGCAACCAACAA 477
Qy 440 tctctgaatacatatcgctggtctctctggttaccatcaccataaactgaacactgcta 499
Db 478 ACAAAACAGCAACCAACCAACAGCAACAGCAACCAACCAACCAACCAACCAACCAACAA 537
Qy 500 aaatctacataacggttaactggaatctataatccgacatcaagaacatccggtgaagtta 559
Db 538 ACAGCAACAGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACGA 597
Qy 560 tcgctaacggtgaatcatcttcaactggaactgacgtgacatcga 602
Db 598 ACAACAAACAAACAGCAACCAACCAACCAACCAACCAACCAACCAACCAACCAACGA 640

RESULT 12
BM278263 546 bp mRNA linear EST 20-DEC-2001
LOCUS AS_tgz_55D05_SKPL Ascaris suum adult male testis germinal zone from
DEFINITION Alan Scott Ascaris suum cDNA clone As_tgz_55D05 5', mRNA sequence.
ACCESSION BM278263
VERSION BM278263.1 GI:17971521
KEYWORDS EST.
SOURCE pig roundworm.
ORGANISM Ascaris suum
Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea
; Ascarididae; Ascaris.
REFERENCE 1 (bases 1 to 546)
AUTHORS Blaxter,M.L., Parkinson,J., Whitton,C., Daub,J., Guiliano,D., Hall
,N., Quayle,M. and Barrell,B.
TITLE Edinburgh University/Sanger Centre Nematode EST Project
JOURNAL Unpublished (2000)
COMMENT Contact: Blaxter ML
Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
3JT, UK.
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk
The library was prepared by Michelle Lizotte-Waniewski for Alan
Scott, Johns Hopkins University Medical School, Baltimore.
Sequencing was performed by the Pathogen Sequencing Unit, Sanger
Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart Barrell). The
sequence contained a PolyA tail (trimmed)
PCR Primers
FORWARD: T3
BACKWARD: T7PL
Plate: 55 row: D column: 05
Seq primer: SKPL
High quality sequence stop: 546.
FEATURES
source 1..546
/organism="Ascaris suum"
/db_xref="taxon:6253"
/clone="As_tgz_55D05"
/location="Edinburgh, UK"
from Alan Scott"
/sex="Male"
/dev_stage="Adult"
/note="Vector: Lambda Zap II; Site_1: EcoRI; Site_2: XhoI;
Library was made from dissected testis germinal zone from
adult male Ascaris suum collected from abattoirs.
Constructed by Michelle Lizotte-Waniewski for Alan Scott,
Johns Hopkins University Medical School, Baltimore, MD."
BASE COUNT 206 a 151 c 68 g 121 t
ORIGIN

```

```

Query Match 3.8%; Score 51.6; DB 10; Length 546;
Best Local Similarity 48.3%; Pred. No. 0.0083;
Matches 175; Conservative 0; Mismatches 184; Indels 3; Gaps 1;

Qy 289 tacatccacaatgaatacacaccatcatcgaactctgactctgatacatcaacggtgaagacaatct 408
Db 41 TCCAACCTGGAATAGCAACAGCTTTTTCATAGCAATTTCAACCAACCAACCCCTTTCACCAAC 100
Qy 349 tccatccgcggtgaacggtatcatctggaactctgactctgatacatcaacggtgaagacaatct 408
Db 101 AGCAACTTCAACCAACCAACTTCAACAGCAATTTCAACCAACCAACCAACTTCAACAGCAACT-- 158
Qy 409 gtattctcgaatacaacatccggtgaagacatctctgaaatcacatcaatcggtgttttc 468
Db 159 -TCACCAACCAACTTCAACAGCAATTTCAACCAACCAACCAACTTCAACAGCAACTTTCAGCAGC 217
Qy 469 gttaccatcaccataacacctgaacaaatgctaaaatctacatcaacggtgaagacaatct 528
Db 278 CTCAACCAACCTCAATTAACCAACCTCAACCAATTTCAACCAACCAACCTCACTAAGCAACCAAC 337
Qy 589 gacggtgacatcgatcgatcccggttctatctggtgatgaatactctccatcttcaacacc 648
Db 338 AACCTTAGCAACCTCAACCAACCAACTTCAACCAACCAACCAACTTCAACAGCAACTTTCAC 397
Qy 649 ga 650
Db 398 AA 399

RESULT 13
BM278164 519 bp mRNA linear EST 20-DEC-2001
LOCUS AS_tgz_54A11_SKPL Ascaris suum adult male testis germinal zone from
DEFINITION Alan Scott Ascaris suum cDNA clone As_tgz_54A11 5', mRNA sequence.
ACCESSION BM278164
VERSION BM278164.1 GI:17971422
KEYWORDS EST.
SOURCE pig roundworm.
ORGANISM Ascaris suum
Eukaryota; Metazoa; Nematoda; Chromadorea; Ascaridida; Ascaridoidea
; Ascarididae; Ascaris.
REFERENCE 1 (bases 1 to 519)
AUTHORS Blaxter,M.L., Parkinson,J., Whitton,C., Daub,J., Guiliano,D., Hall
,N., Quayle,M. and Barrell,B.
TITLE Edinburgh University/Sanger Centre Nematode EST Project
JOURNAL Unpublished (2000)
COMMENT Contact: Blaxter ML
Institute of Cell, Animal and Population Biology
University of Edinburgh
Ashworth Labs, King's Buildings, West Mains Road, Edinburgh, EH9
3JT, UK.
Tel: +44 131 650 6760
Fax: +44 131 670 5450
Email: mark.blaxter@ed.ac.uk
The library was prepared by Michelle Lizotte-Waniewski for Alan
Scott, Johns Hopkins University Medical School, Baltimore.
Sequencing was performed by the Pathogen Sequencing Unit, Sanger
Centre, Cambridge, UK (Neil Hall, Mike Quail & Bart Barrell). The
sequence contained a PolyA tail (trimmed)
PCR Primers
FORWARD: T3
BACKWARD: T7PL
Plate: 54 row: A column: 11
Seq primer: SKPL
High quality sequence stop: 519.
FEATURES
source 1..519
/location/Qualifiers

```



```
source
1. .522
/organism="Ascaris suum"
/db_xref="taxon:6253"
/clone="As_tgz_64B11"
/clone_lib="Ascaris suum adult male testis germinal zone
from Alan Scott"
/sex="Male"
/dev_stage="Adult"
/notes="Vector: Lambda Zap II; Site_1: EcoRI; Site_2: XhoI;
Library was made from dissected testis germinal zone from
adult male Ascaris suum collected from abbatolirs.
Constructed by Michelle Lizotte-Waniewski for Alan Scott,
Johns Hopkins University Medical School, Baltimore, MD."
BASE COUNT      198 a      143 c      113 t
ORIGIN

Query Match      3.7%; Score 49.2; DB 10; Length 522;
Best Local Similarity 48.3%; Pred. No. 0.035;
Matches 169; Conservative 0; Mismatches 178; Indels 3; Gaps 1;

Qy 195 gaatcagaacatcttcacactccgtattccctggaacttctctgtttcccttcggttcg 254
    |||| |||| || |||| || |||| || || || || || || || || || || || ||
Db 67 GAATAGCAACAGCAACTTCAACAGCAATTTCAACAACAATTTCAACACAGCAACTT 126

Qy 255 tatccgaaatcacagacgacggtatccagaattaccatccacaatgaatacacaccatcat 314
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 127 CAACAACAACCTTCAACAGCAACTTCAACAGCAACTTCAACAACAACCTTCAACAGCAATTT 186

Qy 315 caactgcatgaagaataactctgtgtggaagatctccatccgcggtaaccggtatcatctg 374
    |||| || || || || || || || || || || || || || || || || || || ||
Db 187 CAACAACAGCAACTTCAACAGCAATTTCAACAGCAACTTCAACAGCAATTTACCTCAACAA 246

Qy 375 gactctgatgatatacaacggttaa--gaccaaatctgtattcttcgaatacaacatccg 431
    |||| || || || || || || || || || || || || || || || || || || ||
Db 247 CTCCTTAAACAACCTCAACTATATACTTGAACAATCTCAACAACCTCAATAACAACCTCAA 306

Qy 432 tgaagacatctctgaatacacatcgcgtgttcttcgttaccatccccaataacacctgaa 491
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 307 CAATCTCAACAACAACCTCAGTAACCTCAACAACAACCTTAGCAACCTCAGCAACAGTAA 366

Qy 492 caatgctaaaaatctacatcaacggttaactggaatcttaataacogacatcaa 541
    | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 367 CTTCAACAACAGCAACTTCAACAGCAATTTCAACAATTTTAGCAATAGCA 416
```

Search completed: September 2, 2002, 16:32:26
Job time: 5463 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2002, 16:01:38 ; Search time 199.71 Seconds
(without alignments)
11528.628 Million cell updates/sec

Title: US-09-910-186a-7

Perfect score: 1341

Sequence: 1 gaattcacgatgccacaa.....ggaccgaataagtaagaattc 1341

oring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_032802.*
1: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/hold-geneseg/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1341	100.0	1341	21	AAZ87213
2	1341	100.0	1341	22	AAA54485
3	1271.4	94.8	1347	22	AAA54589
4	741.2	55.3	1547	19	AAV30581
5	739.6	55.2	3876	19	AAV30579
6	662.8	49.4	1526	19	AAV30580
7	522.4	39.0	3509	19	AAV26289
8	426.8	31.8	1368	22	AAA54491
9	418.6	31.2	1347	21	AAZ87217

10	263	19.6	1323	22	AAA54483	Botulism toxin hea
11	263	19.6	1326	22	AAA54484	Botulism toxin hea
12	263	19.6	1330	17	AAT29245	Type A neurotoxin
13	263	19.6	1330	19	AAV30571	Clostridium botuli
14	263	19.6	1332	22	AAA54482	Botulism toxin hea
15	263	19.6	1338	21	AAZ87212	DNA encoding synth
16	263	19.6	1351	19	AAV30576	Clostridium botuli
17	263	19.6	1402	17	AAT29246	Type A neurotoxin
18	263	19.6	1402	19	AAV30572	Clostridium botuli
19	242.6	18.1	1313	18	AAT48101	Immunogenic type F
20	242.6	18.1	1314	22	AAA54499	Botulism toxin hea
21	242.6	18.1	1317	22	AAA54490	Botulism toxin hea
22	241.6	18.0	1317	21	AAZ87216	DNA encoding synth
23	227.2	16.9	1332	22	AAA54588	Sequence encoding
24	221.6	16.5	702	21	AAZ87221	DNA encoding BONTA
25	218.6	16.3	1535	19	AAV30596	Clostridium botuli
26	182.2	13.6	1374	22	AAA54487	Botulism toxin hea
27	160	11.9	1278	22	AAA54489	Botulism toxin hea
28	160	11.9	1400	21	AAZ87215	DNA encoding synth
29	160	11.9	1400	22	AAA54488	Botulism toxin hea
30	134.2	10.0	3712	16	AAQ97490	Plasmid pHTAL. N
31	131.4	9.8	1359	12	AAQ12121	Synthetic tetanus
32	129.4	9.6	4366	16	AAQ85424	Plasmid pTECH3-P28
33	129.2	9.6	3754	15	AAQ57880	Intermediate plasm
34	129.2	9.6	3754	16	AAQ85420	Plasmid pTECH1. S
35	129.2	9.6	3769	15	AAQ57881	Intermediate plasm
36	129.2	9.6	3769	16	AAQ97492	Plasmid pTECH2. N
37	129.2	9.6	4377	16	AAQ85421	Plasmid pTECH1-P28
38	123	9.2	1317	21	AAZ87220	DNA encoding nativ
39	123	9.2	1546	19	AAV30575	Clostridium botuli
40	123	9.2	2532	21	AAZ87218	DNA encoding nativ
41	123	9.2	3891	17	AAT29244	C. botulinum type
42	123	9.2	4835	21	AAC64582	BONT/A neurotoxin
43	120.4	9.0	1371	22	AAZ87214	DNA encoding synth
44	120.4	9.0	1371	22	AAA54486	Botulism toxin hea
45	103.2	7.7	1766	16	AAQ97491	Plasmid pTECH2 + h

ALIGNMENTS

RESULT 1

AAZ87213

ID AAZ87213 standard; DNA; 1341 Bp.

XX AC AAZ87213;

XX DT 08-MAY-2000 (first entry)

XX DE DNA encoding synthetic BoNT serotype B (BoNTB) Hc fragment.

XX DE Botulinum neurotoxin; heavy chain; BoNT; serotype B;

XX KW C-terminal fragment; Venezuelan equine encephalitis virus replicon;

XX KW VEG; botulism; vaccine; diagnosis; drug screening; ds.

XX OS Clostridium botulinum.

XX OS Synthetic.

XX FH Key Location/Qualifiers

XX FT CDS 10..1333

XX FT /*tag= a

XX FT /product= "Synthetic botulinum neurotoxin serotype B (BoNTB) heavy chain C-terminal fragment (Hc)"

XX PN WO200002524-A2.

XX PD 20-JAN-2000.

XX PE 09-JUL-1999; 99WO-US15570.

XX PR 10-JUL-1998; 98US-0092416.

XX PR 12-MAY-1999; 99US-0133870.

XX

1104 aattcctaataaagatgatagtagttagaagaagattatataatctagatttt 1163
970 ttcaacctgaatcagaatggcgtgtatataccactacaagtagtacttcaagaagaagaaga 1029
1164 tttaattcaaatcaagagtgagatgatatataccctataaatttttaagaagaagaaga 1223
1030 aagcttttccctggctccgctatctgtatcccgacgaactctacaacaccatccagatcaa 1089
1224 aaattgttttagctccatgaagtgtatctgatgagttttcaatactatacaataaaa 1283
1090 gaatcagcaaacagccactactcttggccagctgtgttcaagaagaatgaagaatct 1149
1284 gaatatgatgaacagccacatatagttgtcagttgctgttttaaaaaagatgaagaagt 1343
1150 actgacgaataatcgtctgatcgtgtatcaacacggtttctcgaatcgtgtatctcga 1209
1344 actgatgagatgaggtgatggtgtatcattcgtttctcgaatcgtgtatctcgaatcgt 1403
1210 gaatacaagactactcttctgcatctccaaatggtacctgaaggaagttaaacgcgaaccg 1269
1404 gagtataagattattttgtataagtaaatggtacttaaaagaggttaaaaggaaccca 1463
1270 tacaacctgaactgggttgcattggcagttctatcccgaaagacgaaggttggaccgaa 1329
1464 tataatttaaaattggatgtaattggcagttattctctaaagatgaaggtggactgaa 1523
1330 ta 1331
1524 ta 1525

RESULT 5
AAV30579
ID AAV30579 standard; DNA; 3876 BP.
XX AAV30579;
AC AAV30579;
XX 07-DEC-1998 (first entry)
XX Clostridium botulinum type B toxin gene from Danish strain.
XX Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW botulism; BotB; ds.
XX Clostridium botulinum serotype B Danish strain.
XX W09808540-A1.
XX 05-MAR-1998.
XX 28-AUG-1997; 97WO-US15394.
XX 28-AUG-1996; 96US-0704159.
XX (OPHI-) OPHIDIAN PHARM INC.
XX Thalley BS, Williams JA;
XX WPI; 1998-230234/20.
XX P-PSDB; AAW68392.
XX Host cell containing recombinant expression vector encoding
PT Clostridium botulinum type B or E toxin - useful to treat humans
PT and other animals at risk of intoxication with clostridial toxin
XX Example 35; Page 291-296; 428pp; English.
XX This is the coding region of the Clostridium botulinum serotype B
CC (Danish strain) toxin gene that codes for a 1291-amino acid
CC polypeptide (see AAW68392). The C fragment (see AAW68394) of the
CC B toxin has been expressed as histidine-tagged protein in Escherichia
CC coli host cells. The invention relates to C. botulinum recombinant

toxin polypeptides. Methods are provided which allow for the
CC isolation of soluble recombinant proteins free of significant
CC endotoxin contamination.. Preferred hosts for production of the
CC recombinant proteins are E. coli, insect cells and yeast cells.
CC The recombinant proteins are used as immunogens for the production
CC of vaccines and antitoxins that are useful in the treatment of
CC humans and animals at risk of intoxication with clostridial toxin.
XX
SQ Sequence 3876 BP; 1612 A; 370 C; 617 G; 1277 T; 0 other;
Query Match 55.2%; Score 739.6; DB 19; Length 3876;
Best Local Similarity 72.5%; Pred. No. 4.2e-194;
Matches 958; Conservative 0; Mismatches 364; Indels 0; Gaps 0;
Qy 10 atggcccaacaataacaaattccgaaatcccgaaacataatcctcgaaacactcggttacaaa 69
Db 2554 atgtttaataataatagcgaaattttaataataatcttaaaatttaagataata 2613
Qy 70 gacacaatctgatcgatctgtctggttacggtgctaaagtgaagtatacgcggtgtt 129
Db 2614 gataataattcaatagattttacagatatggggcaaaaggtataggtatgatggagtc 2673
Qy 130 gaactgaatgacaagaaccagtttcaaaactcctcttcgcgttaactcttaagatccgtgtt 189
Db 2674 gagcttaatgataaaaaatcaatttaaaactagtttcagcaaatagtaagattagagt 2733
Qy 190 actcagaatcagaacatcattctcaaaactcgtattctcgagacttctgttctctctgg 249
Db 2734 actcaaatcagaatcattcttaaatagttgttctcgtgatttagcgttagctttgg 2793
Qy 250 attcgtatccgaaatacaagaacgagctatccagaatattacatccacaaatgaatacacc 309
Db 2794 ataagaatacctaaataataagaatgattgtatatacaaaaattattcataaagaataaca 2853
Qy 310 atcatcaactcgtatgaagaataaactcgtggttgggaagtctccatccgcggttaacogtgc 369
Db 2854 ataattaatgtatgaaaaataaattcgggtcggaataatctctatttaggggtaagagata 2913
Qy 370 atctggactcgtatcgtatcaacggttaagacaaactctattcttcgtatccgaatacaac 429
Db 2914 ataggacttcaattgtatataaattggaaaacaaactcgtatttttggataataacata 2973
Qy 430 cgtgaagacatctctgaatacatcaatcgtggttcttcgttaccatccacaaataacctg 489
Db 2974 agagaagatatcatagatataataatagattgtttttgtaactattactactaataattg 3033
Qy 490 acaaatgctaaaaatctacatcaacggttaaaactggaatcattatcgcacatcaaaagacatc 549
Db 3034 aataacgctaaaaatttatattaatggttaagctagaatacaaaatcacagatatataagata 3093
Qy 550 cgtgaagttatcgtacggtgaaatcattctcaaaactggacgtgacatcgatcggtacc 609
Db 3094 agagaagtattgctaatgggaaataataatttaaatgagattggtatagatagagaaca 3153
Qy 610 cagttcatctggtgaaataactctccatcttcaacacccgaactgtctcagtcacaaatc 669
Db 3154 caattttattggtgaaataatttcagttatttttaatacgggaatgaatcaatcaataatt 3213
Qy 670 gaagaacggttacaagatccagttcttactccgaatactctgaaagactctctgggtaactcg 729
Db 3214 gaagaagagatataaaaattcaatcatatagcgaataatttaaaagatttttggggaactct 3273
Qy 730 ctgattgtacaacaagaataactatattgttcaatcgtgtgtaacaagaactcttcaatacaa 789
Db 3274 ttaattgacaataagaataattatgttttaattcgggggaataaaaaattctataataa 3333
Qy 790 ctgaagaagaactctccggttgggaaactcgtgactcgttccaaatacaaacacagaactct 849
Db 3334 ctgaagaagaattccactgtggtgaaatttcaacacgtagcaaaataataaataaataatct 3393
Qy 850 aaatacatcaactaccgcgacctgtacatcgtgtgaaagtttcatcctcgtcgcaaatct 909


```
Db 924 ttaagtataaagaattattatattgtttaatcggggaataaaaaattcatattaaa 983
QY 790 ctgaagaagaactctccgtgttggtgaatacctgactcgttccaaatacaaccagaactct 849
Db 984 cagtgaaagattcatctgttagtggaattataataacgtagcaaatataatcagaattcc 1043
QY 850 aaatacatcaactaccgcgacactgtacatcggtgaaagttcatcatccgtcgcaaatct 909
Db 1044 aattataaattataagaatttataattggaagaaaaatttattataagaagagagtc 1103
QY 910 aaactctagtcacaaatgatgacatcgtaacgtacgttaaaagaagacatactacctggacttc 969
Db 1104 aattctcaatctataaataatgatatagttagtaaaagaagattatatacatctagatttg 1163
QY 970 ttcaacctgaatcaggaaatggcgtgtatatacaactcaagaacttcaagaagaagaagaa 1029
Db 1164 gtacttcaccatgaagagtggaagagtataatgctataataattttaaggaacagagaagaa 1223
QY 1030 aagcttttctcgtccgactctctgattccgcagcaactctcaaacaccatccagatcaaaa 1089
Db 1224 aaatgttttattataatgaatgaatttataagaactatagaataaaaa 1283
QY 1090 gaatacgaacagacgcgacctactcttgcagctgctgttccaaagaagaagatgaatct 1149
Db 1284 gaatatgatgaacagccatcatatagttgtcagttgcttttttaaaaaagatgaagaagt 1343
QY 1150 actgacgaatacgtctgctgctgctacacccgtttctacgaatctggtatctatttcgaa 1209
Db 1344 actgagatagaggtggtggtggtatctctcttctacgaatctggtggttttactgtaaa 1403
QY 1210 gaatacaagactactctgcatctccaaatggtacctgaaggaagtttaacgcgaacccg 1269
Db 1404 aagtataaagattattttgtataagtaaatggtacttaaaagaggttaaaaggaaccca 1463
QY 1270 tacaacctgaactgggttgcgaattggcagttccatcccgaaagacgaaggttggaccgaa 1329
Db 1464 tataagtcgaattgggtgtaattggcagtttattctctaaagatgaaggggtggaactgaa 1523
QY 1330 ta 1331
Db 1524 ta 1525
```

RESULT 7

AAV26289
ID AAV26289 standard; DNA; 3509 BP.

XX AC AAV26289;

XX 27-JUL-1998 (first entry)

XX Recombinant botulinum neurotoxin type B LH728/B encoding DNA.

XX Botulinum; recombinant; Clostridium botulinum; neurotoxin;
KW immunogen; detection; tetanus; non-toxic; toxin; ds.

XX Synthetic.

OS Clostridium botulinum.

XX FH Key Location/Qualifiers
FT CDS 1..3509

FT FT /*tag= a
FT FT /product= "LH728/B"
FT FT /note= "no stop codon given"

XX WO9807864-A1.

XX PN 26-FEB-1998.

XX 22-AUG-1997; 97WO-GB02273.

XX 13-DEC-1996; 96GB-0025996.

PR 23-AUG-1996; 96GB-0017671.
XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.
PA (SPEY-) SPEYWOOD LAB LTD.

XX PI Foster KA, Quinn CP, Shone CC;

XX WPI: 1998-169168/15.
XX P-PSDB; AAW56017.

PT Recombinant neurotoxin polypeptides - used to develop therapeutic
PT agents, immunogens or as non-toxic standards for the detection of
XX neurotoxins

PS Disclosure; Page 87-91; 137pp; English.

XX The present sequence encodes a recombinant neurotoxin protein from
CC the present invention. The present invention describes recombinant
CC neurotoxin proteins which comprise a first and second domain, where
CC the first domain is adapted to cleave one or more vesicle or
CC plasma-membrane associated proteins essential to exocytosis, and where
CC the second domain is adapted: (a) to translocate the protein into a
CC cell; (b) to increase the solubility of the protein compared to the
CC solubility of the first domain on its own, or (c) both to translocate
CC the protein into a cell and to increase the solubility of the protein
CC compared to the solubility of the first domain on its own, the protein
CC being free of clostridial neurotoxin (CN) and free of CN precursor that
CC can be converted into toxin by proteolytic action. The recombinant
CC proteins can be used as therapeutic agents for targeting cells
CC expressing a relevant substrate. The products can also be used as
CC immunogens and as non-toxic standards for the assessment and development
CC of in vitro assays for the detection of functional botulinum or tetanus
CC neurotoxins either in foodstuffs or in environmental samples.

XX Sequence 3509 BP; 1468 A; 340 C; 548 G; 1153 T; 0 other;

Query Match 39.0%; Score 522.4; DB 19; Length 3509;
Best Local Similarity 71.7%; Pred. No. 4.9e-134;
Matches 685; Conservative 0; Mismatches 271; Indels 0; Gaps 0;

QY 10 atggccaaacaaatacaattccgaatacctgcgaacataatcctgaacctgcggtacaaa 69
Db 2554 atgtttaataataataatagcgaaatttttaataataattatttataatttaagataaag 2613

QY 70 gacaaactctgatcgtctgtgttacggtgctaaagttgaagtatacacacgtgtt 129
Db 2614 gataataatttaagatttatcaggtatggggcaaggtagaggtatgatggagtc 2673

QY 130 gaactgaatgacaagaaccaggttcaaacctgacctctccgctaaactctaaagtcggtt 189
Db 2674 gagcttaagtataaaatacaatttaataactagttcagcaaaatagtaagtagtg 2733

QY 190 actcagaatcagaacatcatcttcaactccgttattctcgtgactctctgtttctctctgg 249
Db 2734 actcaaaatcagaatcatcatatttaagtgtgttctctgttttttagcgttttgg 2793

QY 250 attcgtatcccgaaatacaagaacgacggttatccagaattatcatccacaataacaccc 309
Db 2794 ataagaatacctaaataaagaatgatggtatcacaaattatataatgaataacata 2853

QY 310 atcatcaactgcgatgaagaataaactctgtgttggaagatctccatccggtgaacggtatc 369
Db 2854 ataataatgtatgaaaaaataattccggtcgtaaaatactatttaggggtaaataggata 2913

QY 370 atctggactctgatcatatcaacggtgaagaccacaatctgtattcttcogaatacaacatc 429
Db 2914 atatggactttaaattgatataaattggaacacccaatcggtatttttttgataatacata 2973

QY 430 cgtgaagacatctctgaatacatcaatcgctggtttcttcgttaccatccaacacgtg 489
Db 2974 agagaagatatcagagatatataatagatggtttttttgttaactattactaataattg 3033

[illegible]

RESULT	8
AAA54491	
ID	AAA54491 standard; DNA; 1368 BP.
XX	
XX	
AC	AAA54491;
XX	
DT	11-APR-2001 (first entry)
XX	
DE	Botulism toxin heavy chain C-terminal coding sequence (serotype G).

WPI; 2001-016048/02.
P-PSDB; AAB04167.

New nucleic acids encoding the carboxy- or amino-terminal portions of the heavy chain of botulinum neurotoxin of serotype A-G, useful as vaccine against botulism

Claim 2; Fig 10a; 73pp; English.

Botulism neurotoxins are translated as a single 150 kDa polypeptide chain and then posttranslationally nicked, forming a dichain consisting of a 100 kDa heavy chain and a 50 kDa light chain which remain linked by a disulfide bond. Nucleic acids encoding the carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy chain of botulinum neurotoxin (BoNT) can be used in recombinant expression vectors and expressed in transformed cells to produce peptide antigens useful for eliciting an immune response to give protective immunity against botulinum neurotoxin, which causes botulism. The nucleic acids are expressible in a recombinant organisms such as *Escherichia coli* or *Pichia pastoris*. The use of recombinant nucleic acids are advantageous since it eliminates the need to culture large quantities of hazardous toxin-producing bacterium. Production yield from the genetically engineered product is also high and cost of production is lower. The nucleic acids can be derived from *Clostridium botulinum* serotypes A-G.

Sequence 1368 BP; 366 A; 445 C; 267 G; 290 T; 0 other:

Query Match 31.8%; Score 426.8; DB 22; Length 1368;
Best Local Similarity 60.5%; Pred. No. 8.7e-108;
Matches 818; Conservative 0; Mismatches 507; Indels 27; Gaps

[illegible]

Db 260 tcaactctgggtccgtaccacaaagtacacaaacaaagacatccagacctacctgcaga 319
Qy 299 atgaatacacatcaactcgtcgaagaataaactcgttggtgaagatccatccgcg 358
Db 320 acgagtaacacatcctccgtgtaacgaagacgtcccggttggaaggtctccatcaagg 379
Qy 359 gtaacggtatcatctggactcgtcgatatacaacggttaagaccacaaatctgtattcttcg 418
Db 380 gaaacggtatcatctggactcgtcgatatacaacggttaagaccacaaatctgtattcttcg 439
Qy 419 aatacaacatccgtgaagacatctcgtgaatacatcaatcgtcgttctcgttacatca 478
Db 440 agtactcatcaagacacatctccgactacatacaacgaaggtctccatcaacatca 499
Qy 479 ccaataaac---ctgaacaatgtaaaatctacatacaacggttaaaactggaatcaataccg 535
Db 500 ccaacgacgctgggtgaacccaacatctacatacaacggttccctcgaagaagtcgcgaga 559
Qy 536 acatcaaaagacatccgtgaagttaacgttaacggtgaatcatcttcaaaactggacggtg 595
Db 560 agatctgaacctggacggtatcaactctccaaagacatcgaactcaagctgatcaact 619
Qy 596 acatcgtatcgtaccagttcatctcgttgatgaataactcttcaatcttcaacacccgaactgt 655
Db 620 gtaccgacaccacaaagttcgtcgttgatgaaggaacttcaacatcttcggtcgtgactga 679
Qy 656 ctcaagtccaatatcgaagaacggtacaagatccagtcagcttactccgaataactgaaagact 715
Db 680 acgcaacgaggtctccctcgttactcgttgatccagtcctccaccacacccctcgaaggact 739
Qy 716 ttctgggttaactcgtcgtgtacaaagaataactatattgttcaatcgtcgtgaacaaga 775
Db 740 ttctgggaaacccactcgttacgacacccagtcactcgttcaacacgggtatgcaga 799
Qy 776 actcttatacaaaactcgaagaagactcctccgttggtggaatcctcgtactcgttccaaat 835
Db 800 acatcatcaatgaagtactcttccaggctccatcgttggtgagaccgccctcgttaccact 859
Qy 836 acaacgaagaactcctaatacatcaactaccgacgactcgtacatcgttgaaagtccatca 895
Db 860 tcaacaaacgcgc-----catcaactaccagaacctgtaccctgggtcgtcgtttcatca 913
Qy 896 tctcgcgaactcctaactc---tcagtcacatcaatgatcagatcgtacgttaagaagact 952
Db 914 tcaagaagcctcccaactcccgtaacatcaacaaacgaacacatcgtccgtgaggtgact 973
Qy 953 acatcactcgtgacttctcaact---gaatcagaagatggcgtgtatatacactcaagat 1009
Db 974 acatcactgaacatcgacaacatctccgacgagtcctcgttactcgttactcgtcgttca 1033
Qy 1010 acttcaagaagaagaagaagcttttctcgtcgtccgactcgtattccgcgaactct 1069
Db 1034 actcaagagatccagaccagctgttctcgttggcccaatacaacgacacacctactctct 1093
Qy 1070 acaacacatccagatcaagaataacacgaacacccagcactcctcgttgcagctcgtct 1129
Db 1094 acgagcgtcgtcagatcaagaagtactcagagaagacacacctacacactcagatccgt 1153
Qy 1130 tcaagaagaatgaagaatctactcagcaaatcgtcgtgatcgttccacacgcttctcag 1189
Db 1154 ggaagaagacacacacacccctcggactcgttggtgatcgttggttaagtctg-----tcaagg 1207
Qy 1190 aatcgtggtatcgtattcgaagaatacaagaactcctcgtatctccaaatgggtaccctga 1249
Db 1208 actacggttaagctcgtgggaacacctcagcaactcctcgttatcctccacggtggtaccctgc 1267
Qy 1250 aggaaggttaaacgcaacccgtacaaactgaaactcgttgggtggaatggcagttcatcccca 1309
Db 1268 gtcgtatctcgaagaacatcaacaaagctcgtcgtgggtgaatcgtcaggttcatccccag 1327
Qy 1310 aagacgaaggttgaccga 1328
Db 1328 tcgacgaggggttgaccga 1346

RESULT 10
AAA54483
ID AAA54483 standard; DNA; 1323 BP.
XX
AC AAA54483;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulism toxin heavy chain C-terminal coding sequence (serotype A).
XX
KW Botulism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection; ds.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
FH Key Location/Qualifiers
FT CDS 13..1317
FT /tag= a
FT /product= H_C peptide fragment
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US12890.
XX
PR 12-MAY-1999; 99US-0133865.
PR 12-MAY-1999; 99US-0133866.
PR 12-MAY-1999; 99US-0133867.
PR 12-MAY-1999; 99US-0133868.
PR 12-MAY-1999; 99US-0133869.
PR 12-MAY-1999; 99US-0133873.
PR 29-JUL-1999; 99US-0146192.
XX
PA (USSA) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI; 2001-016048/02.
XX
P-PSDB; AAB04089.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of
PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as
PT vaccine against botulism
XX
PS Disclosure; Fig 2a; 73pp; English.
XX
CC Botulism neurotoxins are translated as a single 150 kDa polypeptide
CC chain and then posttranslationally nicked, forming a dichain
CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which
CC remain linked by a disulfide bond. Nucleic acids encoding the
CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
CC chain of botulinum neurotoxin (BoNT) can be used in recombinant
CC expression vectors and expressed in transformed cells to produce
CC peptide antigens useful for eliciting an immune response to give
CC protective immunity against botulinum neurotoxin, which causes
CC botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use
CC of recombinant nucleic acids are advantageous since it eliminates
CC the need to culture large quantities of hazardous toxin-producing
CC bacterium. Production yield from the genetically engineered product
CC is also high and cost of production is lower. The nucleic acids can
CC be derived from Clostridium botulinum serotypes A-G.
XX
SQ Sequence 1323 BP; 404 A; 334 C; 242 G; 343 T; 0 other;

Query Match 19.6%; Score 263; DB 22; Length 1323;
Best Local Similarity 53.9%; Pred. No. 1.7e-62;

Matches 639; Conservative 0; Mismatches 525; Indels 21; Gaps 4;

```
Qy 15 caacaaatacattccgaatacctggaacataatcatatcatctgaacctggttacaaagacaa 74
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 24 cactgaatacacaagaacatcatcaatactcatctgaacctggttacaaagacaa 83
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 75 caactgatgatctgctggttacggtgctaaagtgaagtatacagcaggtgtt---- 129
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 84 tcacctgatgcacctgtctgcgtacgtgtccaaatacaacatcggttctaagttaactt 143
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 130 -gaactgaatgacaagaacacgttcaactgacctcttccgctaaacttaagatccgtgt 188
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 144 cgatccgatgacaagaatcagatccagctgttcaactgtgaatcttccaaaatcgaagt 203
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 189 tactcagaatacagaacatacttcaactcgttattctcgtgacttctctgttctctgt 248
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 204 tatectgaagaatgctatctatatacactctatgtagcaaaaacttctccacctctctg 263
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 249 gattcgtatccgaatacagaacacggtgtatccagaattatccacatacgaataacac 308
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 264 gatccgtatccgaatacacttcaactcc-----atctctgaaacatgaataacac 314
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 309 catcatcaactgatgaagaataactctgttggaagatctccatccgcggtgaacctgat 368
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 315 catcatcaactgatgaagaataactctgttggaagatctctgaactacggtgaat 374
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 369 catctgactctgatcatatcaacggtgaagacaaatctgtattcttctgaataacacat 428
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 375 catctgactctgaggaactcagggaatacaacagcgtgtgttatccaaatactctca 434
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 429 ccgtgaagacatctctgaatacatcaatcgtctgttctgttaccatcccaataa--- 485
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 435 gatgatacaactctgtacatacaatcgtggtgatctgttaccatcccaataatcg 494
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 486 cctgaacaaatgctaaatctacatacaacggttaactggaatctaataccgacatcaaga 545
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 495 tctgaataactccaaatactacatacaacggtcgtctgatcagaccagaacggtctccaa 554
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 546 catccgtgaagtattcgtcaactgaactcaactctcaactgaacggtgacatcgatcg 605
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 555 tctggtgcaacatccacgcttctataacatacatgttcaactggaacggtgtgogtgacac 614
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 606 taccaggttcatctggtgaaataactctccatctcaacacacgaactgtctcagtcacaa 665
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 615 tcaccgtcatatctggtatcaataacttcaatctgttcgacaaagaactgaacgaaaga 674
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 666 tatagaagaacggttacaagatcaactcttactcgaatacctgaagaactctctgggtcaa 725
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 675 aatcaagaacgtgacgacacacagtcctcaatctgtgtatctgaaagacttctgggtga 734
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 726 tccgctgatgtacaacaaagaatactataatgttcaatgctgtggttaacagaactcttact 785
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 735 ctacctgagtacgacaaacggtactacaatgctgaaatctgtacgataccgaaataacgt 794
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 786 caaactgaagaagaactctccggtgtgtgaaatctcgtgactcgttccaaatacaccagaa 845
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 795 tgacgtcaacaatgtaggtatccggttaccatgacctgaaaggctccggtgttctgt 854
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 846 ctctaaatacatcaactacacggaactgtacatcgttgaaaggcttcaatccggtcgcaa 905
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 855 tatgactaccaacatctacgtgaactctccgtgtacacgtgtgtaccacaaatcatcatcaa 914
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 906 atctaaactcgaatccatcatgatgatcgtacgtacgttaagaagaactacatctacctgga 965
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 915 gaaatagcgtctgtgtaacaaggacaaatacgttctcgcaacaatgatcgtgtatacatcaa 974
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 966 ctcttcaactgaatcaggaaatggcgtgtatacacatcaagaacttctcaagaagaaga 1025
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 975 tgtgttagttaagaacaaagaataccgtctggtctaccatgcttctcaggtgtgtaga 1034
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1026 agaaagacttctggtcgtcgtatctgtattccgacgaactctataaacaccatccagat 1085
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1035 aaagatcttctgctgtgaaatcccgacgttggtaattctgtcaggtagttgtaat 1094
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Qy 1086 caaagaatacagcaaacagccgacctactctgtccagctgtgttcaagaagaatgaaga 1145
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1095 gaaatccaagaacagccagggta---tcaatacaaatgcaaaatgaatctgcaggacaa 1151
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Qy 1146 atctactgacgaaatcggtctgctgacgtatccaccgttctctacga 1190
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1152 caatggaatacgtatcggtttctcatcggtttccaccaggttcaacaa 1196
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
RESULT 11
AAA54484
ID AAA54484 standard; DNA; 1326 BP.
XX
AC AAA54484;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulinism toxin heavy chain C-terminal coding sequence (serotype A).
XX
KW Botulinism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection; ds.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
FH Key Location/Qualifiers
FT CDS 13..1320
FT FT /*tag= a
FT FT /product= H_C peptide fragment
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US12890.
XX
PR 12-MAY-1999; 99US-0133865.
PR 12-MAY-1999; 99US-0133866.
PR 12-MAY-1999; 99US-0133867.
PR 12-MAY-1999; 99US-0133868.
PR 12-MAY-1999; 99US-0133868.
PR 12-MAY-1999; 99US-0133868.
PR 12-MAY-1999; 99US-0133873.
PR 29-JUL-1999; 99US-0146192.
XX
PA (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
DR WPI: 2001-016048/02.
DR P-PSDB; AAB04090.
XX
PT New nucleic acids encoding the carboxy- or amino-terminal portions of
PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as
PT vaccine against botulism
XX
PS Disclosure; Fig 3a; 73pp; English.
XX
CC Botulinism neurotoxins are translated as a single 150 kDa polypeptide
CC chain and then posttranslationally nicked, forming a dichain
CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which
CC remain linked by a disulfide bond. Nucleic acids encoding the
CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
CC chain of botulinum neurotoxin (BoNT) can be used in recombinant
CC expression vectors and expressed in transformed cells to produce
CC peptide antigens useful for eliciting an immune response to give
CC protective immunity against botulinum neurotoxin, which causes
CC botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use
CC of recombinant nucleic acids are advantageous since it eliminates
CC the need to culture large quantities of hazardous toxin-producing
CC bacterium. Production yield from the genetically engineered product
```

CC is also high and cost of production is lower. The nucleic acids can
XX be derived from Clostridium botulinum serotypes A-G.
SQ Sequence 1326 BP; 404 A; 336 C; 243 G; 343 T; 0 other;

Query Match 19.6%; Score 263; DB 22; Length 1326;
Best Local Similarity 53.9%; Pred. No. 1.7e-62;
Matches 639; Conservative 0; Mismatches 525; Indels 21; Gaps 4;

QY 15 caacaaatacaattccgaatacctgaaacataatcatcctgaacactgcgttacaaagacaa 74
DB 27 caatgatacatcaagaacatacatcaatcctcctgaacactgcgttacaaagacaa 86
QY 75 caatcgatcgatctgctggttacggtgctaaagtgaagtgaatagacggtgtt---- 129
DB 87 tcacctgatcgacctgtcgtctacgtctccaaatacaacatcggtttcaaaagttaactt 146
QY 130 -gaactgaatgacaaagacacaggttcaaaactgaccttctccgttaactctaaagatccggt 188
DB 147 cgatcgatcgacaaagacacagatccagctgttcaatctggaatcttccaaatacgaa 206
QY 189 tactgaatcgacaaacatactcttcaactccgttatcttcctggacttctgtttctcttg 248
DB 207 tatcctgagaatgctcatgctatcacactctatgtacgaaacttctccacctctcttg 266
QY 249 gattctatcccgaaatacaagaacgacggtatccagaattatcacaaatgaatacac 308
DB 267 gatcgtatcccgaaatactcaactcc-----atctctctgaacaaatgaatacac 317
QY 309 catcatcaactgcatgaagaataactctggttggagaatctccatccgcggttaacogtat 368
DB 318 catcatcaactgcatggaacaaattctggttggagaatctctgaactacggtgaaat 377
QY 369 catctgagctgcatgatacaacggtgaagacaaatctgtattcttcggaatacaaat 428
DB 378 catctgagctgcatgatacaacggtgaagacaaatctgtattcttcggaatacaaat 437
QY 429 cgtggaagacatctctgaatacatcaatcgctggttcttcgttaaccatcaccaataa--- 485
DB 438 gatgatacaactctgactatacaataatcgctggtgacttctgttaaccatcaccaataatcg 497
QY 486 cctggaacatgcttaaatatcatcaacggtgaactcttaactacccgacatacaaa 545
DB 498 tctgaataactcaaaatctatcaacacgctgctgacgacagaaacccgactctccaa 557
QY 546 catcctggaagtatcgctaaacggtgaataatctctcaactgacggtgacatcgatcg 605
DB 558 tctgggttaacatccacgctcttaataacatcatgttcaaaactggacggttctgtgacac 617
QY 606 taccagttcatctgataaataacttctccatcttcaacacgacgaaactgtctcaagccaa 665
DB 618 tcaccgtcatctggtgataaataacttctcaatctgtctcgacaaagaactgaacgaaaaa 677
QY 666 tatcgaagaacgtgataaagactcttactccgcaatacctgaaagactcttgggtgtaa 725
DB 678 aatcaagaacgtgacgaacacacgctccaatcttggtatctcctgaagactcttgggtga 737
QY 726 tccgctgtagtacaacaaagaataactatagtctcaatgtgtgtaacaaagaactctacat 785
DB 738 ctactcgtgtagtacaacaaacgcttactacatgtgtaactgtgacgacgaaacaaactg 797
QY 786 caaactgaagaagaactctccggtgtgtgtaaaactctgactcgttccaaatacaacagaa 845
DB 798 tgacgtcaacaaatgtaggatccggtgttacaatgtaacctgaaagggtccggtgttctgt 857
QY 846 ctctaataatacaatacctacgacgacctgtacatcggtgaaagtctcatcctcgctcgca 905
DB 858 tatgactaccaacatactacatgaactcttccctgtaccgtggtaccaaattcatcaaa 917
QY 906 atctaactcagtcacataatgatgacatcgtaacgtaagaagacatacatctacctgga 965
DB 918 gaaatacgcgtctggttaacaaagacaataatgttgcgaacaatgatgctgtatatacaaa 977

QY 966 ctcttcaacactgaatcaggaatggtgtatatacaactacaaagtacttcaagaaaga 1025
DB 978 tgtgtagtttaagaacaagaataaccgtctggtaccaatgcttctcaggtgtgtaga 1037
QY 1026 agaaagcttttctgctccgactctgtatctccgacgaaactctacaacacccatccagat 1085
DB 1038 aaagatctgtctgctggaatcccgacgttggtaactctgctcaggtagtgtaat 1097
QY 1086 caaagaatacgaacagacgacacttactcttccagctgctgttcaagaaagatgaaga 1145
DB 1098 gaaatccaagaacgacagggta---tcaatacaaatgcaaatgaatcgtcaggagcaa 1154
QY 1146 atctatgacgaatacgtgtctgacgtatccaccggtttctacga 1190
DB 1155 caatggttaacgatatcgttttcacgtgttccaccaggttcaacaa 1199

RESULT 12

AAT29245

ID AAT29245 standard; DNA; 1330 BP.

XX AC AAT29245;

XX AC AAT29245;

XX DT 07-JUL-1996 (first entry)

XX XX Type A neurotoxin C fragment synthetic gene.

XX DE Toxin; neurotoxin; fusion protein; antitoxin; vaccine; immunogen;

XX KW Clostridium botulinum; ds.

XX OS Synthetic.

XX FH Key Location/Qualifiers

XX FT CDS 1..1317

XX FT /*tag= a

XX PN W09612802-A1.

XX PD 02-MAY-1996.

XX PF 23-OCT-1995; 95WO-US13737.

XX PR 07-JUN-1995; 95US-0480604.

XX PR 24-OCT-1994; 94US-0329154.

XX PR 16-MAR-1995; 95US-0405496.

XX PR 14-APR-1995; 95US-0422711.

XX PA (OPHI-) OPHIDIAN PHARM INC.

XX PI Firca JR, Kink JA, Padhye NV, Stafford DC, Thalley BS;

XX PI Williams JA;

XX DR WPI; 1996-230603/23.

XX DR P-PSDB; AAR95008.

XX PT Fusion proteins comprising non-toxin protein and part of toxin

XX PT useful to form anti-toxins against Clostridium botulinum type A, and

XX PT C. difficile type toxins, and to treat C. difficile intoxication,

XX PT partic. diarrhoea

XX PS Example 22; Page 336-38; 434pp; English.

XX CC A synthetic gene (AAT29245) codes for the heavy chain C fragment

XX CC (AAR95008) of Clostridium botulinum type A neurotoxin (see also

XX CC AAR95010). Codon usage allowing efficient gene expression in Escherichia

XX CC coli was utilised. The gene in vector pALTERBot was used to make

XX CC expression constructs in which fragments of C. difficile toxin A

XX CC repeat domains were expressed as genetic fusions with the C.

XX CC botulin C fragment and expressed in E. coli.

XX SQ Sequence 1330 BP; 400 A; 339 C; 246 G; 345 T; 0 other;

Query Match				19.6%;	Score 263;	DB 17;	Length 1330;
Best Local Similarity				53.9%;	Pred. No. 1.7e-62;		
Matches 639; Conservative				0;	Mismatches 525;	Indels	21; Gaps
QY	15	caacaaatacaattccgaaatcctgaaacataatcatctggaacctcgcttacaagaacaa	74				
Db	24	cactgaatacatcaagaacatcatcaatcaccctccatcttgcgaacctgcgtacgaatccaa	83				
QY	75	caatctgatcgatctgcttggttacgggtgctaaagtgaagtatacgaacgggtgtt----	129				
Db	84	tcacctgatcgactgctgcgtacgcttccaaaatcaacatcggttctaaagtttaactt	143				
QY	130	-gaactgaatgacaagaacacaggttcaaacctgaccttcttcgctaacttaagatccgtgt	188				
Db	144	cgatccgatcgacaagaatcagatccagctgttcaatcggaatctccaaaatcgaagt	203				
QY	189	tactcagaatcagaacatcatcttcaactccgtattcctggactctctgtttcctcttg	248				
Db	204	tatcttgaagaatgctatgatacaactctatgacgaaaacttctccactcctcttg	263				
QY	249	gattcgtatcccgaatacaagaacgacgggtatccagaattatcatccaatgaatacac	308				
Db	264	gatccgtatcccgaatacttcaactcc-----atctctggaacaatgaatacac	314				
QY	309	catcatcaactgcatgaagaataaactctggttgggaagtctccatcccgcggttaaccgtat	368				
Db	315	catcatcaactgcatggaagaaacaattctggttgggaagtatctctggaactcggtgaaat	374				
QY	369	catctggactctgatgatatacaacggtaagaccaaaactgtattcttcgaatacaaat	428				
Db	375	catctggactctgaggaacactcaggaaatacaacagcggtgtgttatccaaatctctca	434				
QY	429	ccgtgaagacatcttgaatacatacaatcgctgggtcttcctcggttacatcaaccaataa	485				
Db	435	gatgatacaactctcgatacatacaatcgctggatcttcgtttacatccacaacaatcg	494				
QY	486	cctggaacaatgctaaaatctacatacaacggtaaacctggaatctaataccgcatacaaga	545				
Db	495	tcfgaataaactccaaaatctacatacaacggcgctgctgacgcagaaacggatctccaa	554				
QY	546	catccgtgaagtattatcgtaaacggtaaatcatcttcaaacctggacgggtgacatcgatcg	605				
Db	555	tcfgggtaacatcccgctctctaataaacatcatgttcaaacctggacgggtgtctgtaaac	614				
QY	606	taccaggttcaatcgatgataaatacttctccatcttcaacacacgaactgtctcagttccaa	665				
Db	615	tcaccgctacatctggatcaaatcttcaattctgctgacaaaagaactgaaacgaaaaaga	674				
QY	666	tatcgaagaacggtatacaagatccagttcttacttccgaaatacctgaaagacttctgggttaa	725				
Db	675	aatcaagacctgtacgacaacacagtcctcaattctggttatcctgaaagacttctgggttga	734				
QY	726	tcgcgtgatgtacacaagaataataatgttccaatgctggtgaacagaactcttcat	785				
Db	735	ctacctgcagtaacaaaacggtactacaatgctggaatctgtacgatccgaaacaatacgt	794				
QY	786	caaacgtgaagaagactctccggttggtgaaatcctgactcgttccaaatacaaccagaa	845				
Db	795	tgcgttcaacaatgtaggttatccggggttaacatgtaacctgaaaggtccgctggttctctg	854				
QY	846	ctctaaatacatcaactaccggaactgtacatcggttgaaagttcatcatccgtcgcaa	905				
Db	855	tatgataccaacatcatcctgaactcttcccctgtaccggtggtaccaaatctcatcaaca	914				
QY	906	atctaactctcagtcctaataatgatgacatcgtaacgttaagaagactacattaccttggga	965				
Db	915	gaaatcgcgctctggttaacagaagacaatatcgttcgacaacatgatcgtgtatacatcaa	974				
QY	966	cttcttcaaccttgatccaggaatgcggtgatacaacctacaaagtacttcaagaagaaga	1025				
Db	975	tgttgtagttgaagaagaataaccgtctggctaccacgaatctctcagcgctggtgtgaga	1034				

	Qy	1026	aagaaacgttttctggtccgatccttgatttcgcacgaactctacaaccatccagat	1085
	Db	1035	aagatcttgtctgctggaaatcccgcagtgttgtaatactgtctcgtagttgtaat	1094
	Qy	1086	caagaatacacgacgaacaggacctactcttcgccagctgctgttccaagaagatgaaga	1145
	Db	1095	gaatatcaaagacgaccaggcta---tctaatcaaatgcaaaatgaatctgcaggacaa	1151
	Qy	1146	atctactgacgaatacggctcgtgatcggttatccacgcgtttcttacga	1190
	Db	1152	caatggtaaagatacgtgttcacgtgttcaccacgittcaacaa	1196
	RESULT	13		
	AAV30571	ID	AAV30571 standard; DNA; 1330 BP.	
	XX AC AAV30571;			
	XX DT 07-DEC-1998	(first entry)		
	XX DE Clostridium botulinum toxin A fragment C gene in pAlterBot.			
	XX KW Antitoxin; vaccine; neurotoxin; toxin A; intoxication; immunogen;			
	XX KW botulism; ds.			
	XX OS Clostridium botulinum serotype A.			
	XX FH Key Location/Qualifiers			
	FT CDS 1..1317			
	FT FT /*tag= a			
	FT misc_difference 1..6			
	FT FT /*tag= b			
	FT FT /note= "PALTER vector-derived nucleotides (encode Met-Ala)"			
	XX PN W09808540-AI.			
	XX PD 05-MAR-1998.			
	XX PF 28-AUG-1997; 97WO-US15394.			
	XX PR 28-AUG-1996; 96US-0704159.			
	XX PA (OPHI-) OPHIDIAN PHARM INC.			
	XX PI Thalley BS, Williams JA;			
	XX PPI WPI; 1998-230234/20.			
	XX DR P-PSDB; AAU68389.			
	XX PT Host cell containing recombinant expression vector encoding			
	PT Clostridium botulinum type B or E toxin - useful to treat humans			
	PT and other animals at risk of intoxication with clostridial toxin			
	XX PS Example 22; Page 262-263; 428pp; English.			
	CC This is the DNA sequence of the Clostridium botulinum serotype A			
	CC toxin C-fragment gene contained in plasmid pAlterBot. Recombinant			
	CC C-fragment proteins have been produced in Escherichia coli as			
	CC fusion proteins with either maltose binding protein or			
	CC Clostridium difficile type A toxin (see AAU68387). The invention			
	CC relates to recombinant proteins derived from C. botulinum toxins.			
	CC Methods are provided which allow for the isolation of soluble			
	CC recombinant proteins free of significant endotoxin contamination.			
	CC Preferred hosts for production of recombinant proteins are E. coli,			
	CC insect cells and yeast cells. The recombinant toxin proteins are			
	CC used as immunogens for the production of vaccines and antitoxins			
	CC that are useful in the treatment of humans and animals at risk of			
	CC intoxication with clostridial toxin.			
	XX SQ Sequence 1330 BP; 400 A; 339 C; 246 G; 345 T; 0 other;			

Query Match	19.6%; Score 263; DB 19; Length 1330;
Best Local Similarity	53.9%; Pred. No. 1.7e-63;
Matches	639; Conservative 0; Mismatches 525; Indels 21; Gaps 4;
QY	15 caacaaataaattccgaaatctgaaacaatatcatctgaacctgcgttacaaagacaa 74
DB	11
DB	24 cactgaatacatcaagaacatacaatactcatctgaacctgaacctgcgtacgaatcaa 83
QY	75 caatctgacgatctgtctgtttaaaggctgcataaagggtgaagtataacgacgggtt- 129
DB	11
DB	84 tcaactgacgactgtctgtctgaactctcaaaatacaaatcggtgtttaaagttaaactt 143
QY	130 -gaactgaatgacaagaaccaggttcaaaactgacctcttcctgaactctaagatcgtgt 188
DB	11
DB	144 cgatccgatcgacaagaatcagatccagctgttcaatctggaatcttccaaaatgaagt 203
QY	189 tactcagaatcagaacatcatcttcaaacctcggtattctcgactctctgtttctcttg 248
DB	11
DB	204 tatctgaagaatgctatctgtatacaactctatgtacgaaaacttctccactctcttg 263
QY	249 gattcgtatccgaaatacaagaacgacggtatccagaattacatccacaatgaatacac 308
DB	11
DB	264 gatccgtatccgaaatacttcaactc- -----atctcttgaaacatgaatacac 314
QY	309 catcataactgcgatgaagaataactcgtgttgaagaatctccatccgcggttaacgctat 368
DB	11
DB	315 catcataactgcgatgaagaataactcgtgttgaagaatctctgaaactcaggtgaaat 374
QY	369 catctgactctgatcgatatacaacggttaagaccaaaactgtattcttcgataacaacat 428
DB	11
DB	375 catctgactctgaggacacatcgagaaatacaacacgctgtgtattcaaatactctca 434
QY	429 ccgtgaagacatcttgaatacatcaatcgcgtgttcttcttaccatcaccaataaa- 485
DB	11
DB	435 gatgatacaactctgtactacatacaatcgtgtgatcttcttaccatcaaccaaatcgt 494
QY	486 cctgaacaatgctaaaatctcatcaacggttaaacgtgaatcctaaccgcatcaaga 545
DB	11
DB	495 tctgaataactccaaatctacatacaacggtctgatcgaccagaaccgatctccaa 554
QY	546 catccgtgaagtattcgttaacggtgaataatcatcttcaaacgtgacggtgacatcgctg 605
DB	11
DB	555 tctgggttaacatccacgcttcttaataacatcatgttcaaacgtgacggttctgtgacac 614
QY	606 taccagttcatctggatgaataacttctccatcttcaaacacggaactgtctcagtccaa 665
DB	11
DB	615 tcaacgtacatctggatcaaatctcaatctctgttcgacaagaactgaacgaaaaaga 674
QY	666 tatcgaagaacggtacaaagtccagcttactccgaaacctgaagactcttgggttaa 725
DB	11
DB	675 aatcaagacgtgtacgacaacacggtccaaattctgtgtatctcgtgaagaacttctgggtga 734
QY	726 tccgctgatgtacacaagaataactatattgttcaactcgtgtgaacaagaactcttcat 785
DB	11
DB	735 ctactcgtgactgacaaaacggtactacatgctgaatctgtacgataccgacaatacgt 794
QY	786 caaactgaagaaagactctcoggtgtgtgtaaatcctgaactcgttccaaatacaaccagaa 845
DB	11
DB	795 tgaactcaaatgtaggtatcccggttactgttactcgtgaaggtccgctggttctgt 854
QY	846 ctctaaatacatcaactacccgacgtctgacatcggtgaagaagtctatctccgtcgaa 905
DB	11
DB	855 tatgactaccaaacatctacctgaaactctccctgtaccggtgttaccaaatctaatcaatcaa 914
QY	906 atctaatctcagttccatcaatgatgatcatcgttacgttaagaagactacatctcctgga 965
DB	11
DB	915 gaatacgcgtctggttaacagaagcaatcgttctcgcaacaatgatcgtgtatacatcaa 974
QY	966 ctcttcaacctgaatcaggaatggctgtgtatcacactcaagaacttctcaagaagaaga 1025
DB	11
DB	975 tgttgtagtttaagacaagaataacgctctggttaccaaatgcttctcaggtcgtgtga 1034

CC this vector results in the production of large amounts of a protein
CC encoded by a sequence cloned into the replicon. The constructs are used
CC to produce vaccines against botulism. The proteins can also be used as
CC diagnostic tools for the diagnosis of botulism. The transformed host
CC cells can be used to analyse the effectiveness of drugs and agents which
CC inhibit toxin effects. The vaccine currently used against botulism is
CC dangerous and expensive to produce, and contains formalin, which is very
CC painful for the recipient. Also, the vaccine is incomplete, in that only
CC 5 of the 7 serotypes are represented in the formulation. The novel
CC vaccine overcomes these problems, as it is easily purified, and
CC available in large quantities. It is also expressed in the lymph nodes
CC for a better immune response. Sequences A4287212-287217 represent
CC synthetic DNA sequences encoding BoNT Hc fragments used in the present
CC invention. These were optimised for codon usage for expression in yeast.
XX

Sequence 1338 BP; 401 A; 342 C; 249 G; 346 T; 0 other;

Query Match 19.6%; Score 263; DB 21; Length 1338;
Best Local Similarity 53.9%; Pred. No. 1.8e-62;
Matches 639; Conservative 0; Mismatches 525; Indels 21; Gaps 4;

Qy	15	caacaaatacaatccgaatacctgaacaataatcatctcgaacacgtggttaacaagaacaa	74
Db	32	cactgaatacatcaagaacatcatcaatcacctccatctgaacctggcgtacgaatccaa	91
Qy	75	caatctgatcgtctgtctgttaccgtgttaagtgaagtatacgaacgtgtt-----	129
Db	92	tcacctgatcgactgtctcgtcgtcgtctccaaatacaacatcgtgttctaagttaactt	151
Qy	130	-gaactgaatgacaaagaccagtgtcaactgacctcttccgtcaacttaagatccgtgt	188
Db	152	cgatccgatcgacaagaatcagatccagctgttcaatctggaatcttccaaaatcgaagt	211
Qy	189	tactcagaatcagaacatcatcttcaactcogtattctcgtgaactctctgttctctctg	248
Db	212	tatctgaagaatgctatcgtatatacaactctatgtcagaacacttctccactcctctg	271
Qy	249	gattcgtatccgaaatacaagaacgacggtatccagaattacatccacaatgaatacac	308
Db	272	gatccgtatcccgaaatacttcaactcc-----atctctctgacaatgaatacac	322
Qy	309	catcatcaactgcatgaagaataactctgttggaaagtatccatcccggtgaaccggtat	368
Db	323	catcatcaactgcatgaagaataactctgttggaaagtatctctgaactacggtgaat	382
Qy	369	catctggactctgatcatatacaggttaagaacaaatctgtattcttctgaatacacat	428
Db	383	catctggactctgagdacactcaggaaatacaaacagcgtgtgtattcaataactctca	442
Qy	429	ccgtgaagacatctctgaatacatcaatcgtgttctctctgttaccatccacaataa---	485
Db	443	gatgatacaactctctgactacatacaatcgtgtgtgtctctctgttaccatccacaatcg	502
Qy	486	cctgaacaatgctaaatctcatcaacggttaactggaatctaatccgcacatcaaga	545
Db	503	tctgaataactccaaatctcatcaacggtgtgtgtgtgtgtgtgtgtgtgtgtgtgtgt	562
Qy	546	catccgtgaagtattcgttaacggtgaatcatcttcaaacgtgacggtgacatcgatcg	605
Db	563	tctgggttaacatccacgcttctcatatcatatgttcaaacgtgacggtgtgtgtgtgt	622
Qy	606	taccagttcatctggatgaataactctctccatcttccaaacggaactgtctcagttcaa	665
Db	623	tcacgctacatctgttgatcaaatctcaatctgttctgacaaagaactgaacgaaaaaga	682
Qy	666	tatcgaagaacggttacaagatccagtttactccgaatacctcgaagacttctgtgggtta	725
Db	683	aatcaaaagacctgtacgaacacggtccaaattctgtgtatctcctgaagacttctgtgtga	742
Qy	726	tccgctgatgtacaagaagaatactatattgttcaatgctggttaacaagaactcttcaat	785
Db	743	ctactcgtgacgacgaacacggtactacatctgtaattctgtacgatccgacaataacgt	802

Search completed: September 2, 2002, 17:08:55
Job time: 4037 sec

Qy	786	caaaactgaagaaagactctccggttgggtgaataactcctgaactcgtttccaaatacaaccagaa	845
Db	803	tgactcaacaatgtaggtagtaccggtttacatgtaccctgaaggtccgctggttctgt	862
Qy	846	ctctaaaatacatcaactaccgcgacctgtacatcgttgaagaattcatcccgctcgcaa	905
Db	863	tatgataccaacatctaccctgaactcttccctgtaccggtgtaccacaaattcatcatcaa	922
Qy	906	atctaactctcagtcocatcaatgatgacatcgtacgttaagaaagactacatctactgga	965
Db	923	gaatacgcgtctgttaacaaggagcaaatatcgttcgcgaacaatgatcgtgtatatacatcaa	982
Qy	966	cttcttcaacctgaatcagggaatggtgtgtatatacacatcacaagtacttcaagaagaaga	1025
Db	983	tggtgtagttaagaacaaagaataccgctcgtgtaccacaaatgcttctcagggtgtgtaga	1042
Qy	1026	agaaaagcttttctcgtcgcgtatctctgattccgcagcgaacatctacaacacatccagat	1085
Db	1043	aaagatctgtctgtctgtggaatcccgagcgttggtaattctgtctcaggttagttgta	1102
Qy	1086	caaagaatacagacgaacagcgcgaactactcttgcagctgtgttcaagaagaagatgaaga	1145
Db	1103	gaatccaagaacgaccagggtgta---tcaactaacaatgcacaaatgaatctgcaggacaa	1159
Qy	1146	atctactgacgaataatcgtctgacgtgtatccacgcttcttctacga	1190
Db	1160	caatggttaacgatcgtgttccatcgtgttccacggttccacggttccacaa	1204

OM of: US-09-910-186a-8 to: GenEmbl:* out_format : pfs

Date: Sep 2, 2002 3:59 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framet-p2n.model -DEV=xlh
-O=/cn2.1/USPTO.spool/US09910186/runat_29082002.134725.15992/app_query.fasta_1.505
-DB=GenEmbl -QFWT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-ALIGN=7.000 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFWT=pfs -NORM-ext -HEAPSIZE=500
-MINLEN=0 -MAXLEN=200000000 -USER=US09910186_@CGN1_L1_4342
-NCPU=6 -ICPU=3 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30
-NO_XLUPY -WAIT -THREADS=1

Search information block:

Query: US-09-910-186a-8

Query length: 440

Database: GenEmbl.*

Database sequences: 1797656

Database length: 187333701

Search time (sec): 1881.260000

score_list:

Sequence	Strd	Orig	ZScore	Escore	Len	Documentation
gb_ba:CLOBOTB	+	2340.00	2577.97	3.1e-135	4041	M81186 Clostridium botulinum n
gb_ba:AF295926	+	2231.00	2458.15	1.5e-128	3876	AF295926 Clostridium botulinum
gb_ba:AF300465	+	2229.00	2455.94	2.0e-128	3876	AF300465 Clostridium botulinum
gb_ba:AF300466	+	2229.00	2455.94	2.0e-128	3876	AF300466 Clostridium botulinum
gb_ba:CBY13630	+	2213.00	2441.57	1.2e-127	11170	Y13630 Clostridium botulinum
gb_ba:AF300467	+	2210.00	2435.01	2.9e-127	3869	AF300467 Clostridium botulinum
gb_ba:AF300468	+	2210.00	2435.01	2.9e-127	3869	AF300468 Clostridium botulinum
gb_ba:AF300469	+	2210.00	2435.01	2.9e-127	3869	AF300469 Clostridium botulinum
gb_ba:CBBOTB	+	2099.00	2312.35	2.0e-120	4051	AF300469 Clostridium botulinum
gb_ba:CBOT242628	+	2097.00	2318.32	9.1e-121	1326	X71343 C.botulinum type B gene
gb_ba:AF69701	+	1667.00	1837.30	5.6e-94	3509	AJ242628 Clostridium botulinum
gb_ba:BD009886	+	1667.00	1837.30	5.6e-94	3509	AF9701 Sequence 19 from Patent
gb_ba:CBNTG	+	1168.00	1287.07	2.5e-63	3937	BD009886 Recombinant toxin fra
gb_ba:CBNTNHTB	+	1103.00	1210.81	4.4e-59	6862	X71462 C.botulinum BONT/G gene
gb_pat:AR0000029	+	831.50	923.61	4.4e-43	1330	X87849 C.botulinum NTHN gene &
gb_pat:AR169140	+	831.50	923.61	4.4e-43	1330	AR000029 Sequence 22 from pate
gb_pat:AX036243	+	831.50	923.61	4.4e-43	1330	AR169140 Sequence 22 from pate
gb_sy:XXU22962	+	831.50	923.61	4.4e-43	1338	AX036243 Sequence 22 from Pate
gb_pat:AR000030	+	831.50	923.23	4.6e-43	1402	U22962 Synthetic botulinum neu
gb_pat:AR169141	+	831.50	923.23	4.6e-43	1402	AR000030 Sequence 25 from pate
gb_pat:AX036246	+	831.50	923.23	4.6e-43	1402	AR169141 Sequence 25 from pate
gb_pat:AR000031	+	831.50	915.75	1.2e-42	3891	AX036246 Sequence 25 from Pate
gb_pat:AR169142	+	831.50	915.75	1.2e-42	3891	AR000031 Sequence 27 from pate
gb_pat:AX036248	+	831.50	915.75	1.2e-42	3891	AR169142 Sequence 27 from pate
gb_ba:CBOTAG	+	831.50	915.03	1.5e-42	4292	AX036248 Sequence 27 from Pate
gb_ba:CBOTAG	+	831.50	915.03	1.5e-42	4292	X52066 Clostridium botulinum B
gb_ba:CBOTNTF	+	831.50	914.16	1.5e-42	4835	X52066 Clostridium botulinum B
gb_ba:CBNTX	+	778.00	856.45	2.4e-39	4073	M30196 C.botulinum neurotoxin
gb_ba:CBNTX	+	777.50	855.91	2.6e-39	4067	X68262 C.barati gene for type
gb_ba:CBNTX	+	746.00	830.88	6.4e-38	1084	X73423 C.botulinum gene for in
gb_pat:AF58945	+	732.00	814.16	5.5e-37	1293	X70817 C.botulinum (strain pro
gb_sy:AF231281	+	732.00	814.13	5.5e-37	1299	A58945 Sequence 5 from Patent
gb_pat:AF58946	+	732.00	814.05	5.6e-37	1313	AF251281 Synthetic construct
gb_ba:CBOTNTF	+	732.00	805.54	1.7e-36	4199	A58946 Sequence 6 from Patent
gb_ba:CBOTNTF	+	732.00	805.54	1.7e-36	4199	X35496 Clostridium botulinum B
gb_ba:CBY13631	+	723.00	790.32	1.2e-35	9325	X81714 C.botulinum bont/f gene
gb_ba:CBOTNTX	+	723.00	795.56	6.0e-36	4234	X713631 Clostridium botulinum B
gb_ba:CBY13631	+	718.00	800.03	3.4e-36	1084	M92906 Clostridium botulinum B
gb_ba:CBY13631	+	718.00	800.03	3.4e-36	1084	X70819 C.botulinum (strain non
gb_ba:CBNTX	+	692.50	762.30	4.3e-34	4030	X70819 C.botulinum (strain non
gb_ba:CBNTX	+	692.00	761.78	4.6e-34	4017	X62683 C.botulinum bontE gene f
gb_ba:CBNTX	+	692.00	761.78	4.6e-34	4017	X62089 C.botulinum gene for bd

gb_pat:AX088262 + 692.00 761.78 4.6e-34 4017 ! AX088262 Sequence 10 from P
gb_ba:AB037704 + 683.00 752.20 1.6e-33 3835 ! AB037704 Clostridium butyri
gb_ba:AB037705 + 683.00 752.20 1.6e-33 3835 ! AB037705 Clostridium butyri
gb_ba:AB037706 + 683.00 752.20 1.6e-33 3835 ! AB037706 Clostridium butyri
gb_ba:AB037707 + 683.00 752.20 1.6e-33 3835 ! AB037707 Clostridium butyri

seq_name: gb_ba:CLOBOTB

seq documentation block:

LOCUS CLOBOTB 4041 bp DNA linear BCT 26-APR-1993
DEFINITION Clostridium botulinum neurotoxin type B (bont) gene, complete cds.
ACCESSION M81186
VERSION M81186.1 GI:144734
KEYWORDS bont gene; neurotoxin type B.
SOURCE Clostridium botulinum DNA.
ORGANISM Clostridium botulinum
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
Clostridium.
REFERENCE 1 (bases 1 to 4041)
AUTHORS Whelan,S.M., Elmore,M.J., Bodsworth,N.J., Brehm,J.K., Atkinson,T.
and Minton,N.P.
TITLE Complete nucleotide sequence of the Clostridium botulinum gene
encoding the type B neurotoxin
JOURNAL Unpublished (1991)
FEATURES
Location/Qualifiers
1..4041
/organism="Clostridium botulinum"
/db_xref="taxon:1491"
57..3932
/gene="bontB"
57..3932
/gene="bontB"
/function="vertebrate neurotoxin"
/codon_start=1
/transl_table=1
/product="neurotoxin type B"
/protein_id="AA23211.1"
/db_xref="GI:144735"

Source

gene

CDS

BASE COUNT 1679 a 383 c 645 g 1334 t

ORIGIN

alignment_scores:

Quality: 2340.00 Length: 440

Ratio: 5.330 Gaps: 0

Percent Similarity: 99.773 Percent Identity: 99.545

alignment_block:

US-09-910-186a-8 x CLOBOTB

Align seg 1/1 to: CLOBOTB from: 1 to: 4041

```
1 MetAlaAsnLysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLe 17
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2610 ATGTTAAATAAATAAATACGAAATTTTAAATAAATATATCTTAAATTT 2659
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
17 uAqGTYrLYsAspAsnLeuIleAspLeuSerGlyTYrGlyAlaLYsV 34
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2660 AAGATATAGGATAAATAATTAATAGATTTATCAGGATATGGGGCAAGG 2709
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
34 alGluValTYrAspGlyValGLuLeuAsnAspLYsAsnGlnPheLYsLeu 50
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2710 TAGAGGTATATGATGGAGTCGAGCTTAATCATAAAAATCAATTTAAATTA 2759
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
51 ThrSerSerAlaAsnSerLYsIleArgValThrGlnAsnGlnAsnIleI 67
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2760 ACTAGTTCAGCAAAATAGTAGATTAGAGTGACTCAAAATCAGAAATATCAT 2809
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
67 ePheAsnSerValPheLeuAspPheSerValSerPheTrpIleArgIleP 84
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2810 ATTAAATAGTGTCTCTTCGATTAGCGTTAGCTTTGGATAAGAATAC 2859
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
84 roLYsTYrLYsAsnAspGlyIleGlnAsnTYrIleHisAsnGluTYrThr 100
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2860 CTAATATATAGAAATGATGGTATACAAATATATATTCATATGAATATACA 2909
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
101 IleIleAsnCysMetLYsAsnAsnSerGlyTrpLYsIleSerIleArgGI 117
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2910 ATAAATTAATGTATGAAATAAATTCGGCTGGAAATATCTATTAGGGG 2959
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
117 yAsnArgIleIleTrpThrLeuIleAspIleAsnGlyLYsThrLYsSerV 134
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
2960 TAATAGGATAATATGACTTTAAATTCATATAAATGCAAAACCAAAATCG 3009
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
134 alPhePheGluTYrAsnIleArgGluAspIleSerGluTYrIleAsnArg 150
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3010 TATTTTGTGAATATAACATAAAGAGAAGATATATCAGAGTATATAAATAGA 3059
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
151 TrpPhePheValThrIleThrAsnAsnLeuAsnAsnAlaLYsIleTYrIL 167
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3060 TGGTTTTTGTAACTATTAATAATATTTCAATAACGCTAAAATTTATAT 3109
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
167 eAsnGlyLYsLeuGluSerAsnThrAspIleLYsAspIleArgGluValI 184
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3110 TAATGGTAAGCTAGAATCAAAATACAGATATTAAGATATAAAGAGAAGTTA 3159
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
184 leAlaAsnGlyGluIleIlePheLYsLeuAspGlyAspIleAspArgThr 200
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3160 TTCTAATGGTGAATAAATAATTTAAATTTAGATGGTGATATAGATAGAACA 3209
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
201 GlnPheIleTrpMetLYsTYrPheSerIlePheAsnThrGluLeuSerGI 217
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3210 CAATTTATTTGGATGAATAATTTTCACTATTTTAAATACGGAATTAAGTCA 3259
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
217 nSerAsnIleGluGluArgTYrLYsIleGlnSerTYrSerGluTYrLeuL 234
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3260 ATCAAAATATTGAAGAAGATATAAAATTTCAATCATATAGCGAATATTTAA 3309
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
234 yAspPheTrpGlyAsnProLeuMetTYrAsnLYsGluTYrTYrMetPhe 250
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3310 AAGATTTTGGGGAATTCCTTTAATGTACAATAAAGAAATATATATATGTTT 3359
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
251 AsnAlaGlyAsnLYsAsnSerTYrIleLYsLeuLYsAspSerProVa 267
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3360 AATGCGGGGAATAAAATTCATATATTAACATAAACAAGATTCACCTGT 3409
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
267 lGlyGluIleLeuThrArgSerLYsTYrAsnGlnAsnSerLYsTYrIleA 284
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3410 AGGTGAATTTTAAACAGTAGCAATATATAATCAAAATTTCTAAATATATA 3459
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
284 snTYrArgAspLeuTYrIleGlyGluLYsPheIleIleArgLYsSer 300
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3460 ATTATAGAGATTTATATATTTGGAGAAAAATTTATATATAAGAGAAAGTCA 3509
||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

```
301 AsnSerGlnSerIleAsnAspIleValArgLYsGluAspTYrIleTY 317
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3510 AATTCCTCAATCTATAAATGATGATAGTTAGAAAAAGACATTATATATA 3559
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
317 rLeuAspPhePheAsnLeuAsnGlnGluTrpArgValTYrThrTYrLYsT 334
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3560 TCTAGATTTTAAATTTAAATCAAGAGTGGAGAGTATATACCTATAAAT 3609
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
334 yPheLYsLYsGluGluLYsLeuPheLeuAlaProIleSerAspSer 350
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3610 ATTTTAAAGAAAGAGGAAGAAAAATTTCTTTTAGCTCCTATAAGTGATCT 3659
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
351 AspGluLeuTYrAsnThrIleGlnIleLYsGluTYrAspGluGlnProth 367
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3660 GATGAGTTTTACAATACTATACAAATAAAGAAATATGATGAACAGCCAAC 3709
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
367 rTYrSerCysGlnLeuPheLYsLYsAspGluGluSerThrAspGluI 384
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3710 ATATAGTTCTCAGTCTCTTTTAAAAAAGATGAAGAAAGTACTGATGAGA 3759
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
384 leGlyLeuIleGlyIleHisArgPheTYrGluSerGlyIleValPheGlu 400
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3760 TAGGATTGATTGGTATTCATCGTTTCTACGAATCTGGAATTTGATTGAA 3809
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
401 GluTYrLYsAspTYrPheCysIleSerLYsTYrTYrLeuLYsGluVally 417
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3810 GAGTATAAAGATTTATTTTGTATAAGTAAATGGTACTTAAAAGAGGTAAA 3859
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
417 sArgLYsProTYrAsnLeuLYsLeuGlyCysAsnTrpGlnPheIleProL 434
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3860 AAGGAAACCATATAATTTAAATTTGGGATGTAATTTGGCAGTTTATTCCTA 3909
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
434 yAspGluGlyTrpThrGlu 440
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
3910 AAGATGAAGGGTGGACTGAA 3929
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

seq_name: gb_ba:AF295926
seq_documentation_block:
LOCUS AF295926 3876 bp DNA linear BCT 02-SEP-2001
DEFINITION Clostridium botulinum neurotoxin type B gene, complete cds.
ACCESSION AF295926
VERSION AF295926.1 GI:15419707
KEYWORDS
SOURCE Clostridium botulinum.
ORGANISM Clostridium botulinum
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
Clostridium.
REFERENCE
1 (bases 1 to 3876)
AUTHORS Kirma,N., Ferreira,J.L. and Baumstark,B.R.
TITLE Characterization of six type A strains of Clostridium botulinum
that contain type B toxin gene sequences
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 3876)
AUTHORS Kirma,N., Ferreira,J.L. and Baumstark,B.R.
TITLE Direct Submission
JOURNAL Submitted (14-AUG-2000) Department of Biology, Georgia State
University, P.O. Box 4010, Atlanta, GA 30302-4010, USA
FEATURES
source
1. .3876
/organism="Clostridium botulinum"
/isolate="1436"
/db_xref="taxon:1491"
/country="USA: Utah"
/note="isolated from stool sample in 1977
type: AB"
1. .3876
/codon_start=1
/transl_table=1
/product="neurotoxin type B"
/protein_id="AAK97132.1"
/db_xref="GI:15419708"

CDS
```

```
/translation="MSVTNNFNNDPIDNDNIIMMEPPFARGTGRYYKAFKTIIDRIW
IPERYTEGYPEDENKSGGIFNDVCEYIDPDYNTNDKKNIFLQTMIKLFNRKISK
PIGKELMIINGIPYLGDRRVPLPEEFNTASVTNKLISNPGVEVERKGFIFANLII
FGPGVPLNENETIDIGIONHFAREGFGIMQKFCPEVSVFNNOVKGASIFNRR
GYFDPALILMHELHLVGLYKIVNDLPVIVNEKKFFMQSDIAQEELYTFGGQD
PSIISPTDKSIYDKVLQNFGRGIVDRLNKLVLCISDPNININIKYKFKDYKFEVDS
EGKYSIDVESDEKLYKSLMFQGTETNIAENYKIKTRASYFSDSLPPVKIKLLDNEIY
TIEGFNISDKMEKEYRGONKAIKQAYEISKEHLAVYIQMCYSKAPGICIDVD
NEDLFFADKNSFSDLSKNERIAYNTQNNYIDNDFSINELIDTDLTSKTELPSENT
ESLTDENVVPEYKOPAKIKFTDENTIFOYLYSOTPLDIRDLSLTSFDDALLFS
NKVYSFSDMYLKWANKVVEAGLFAGWVKQIVDDPVIENKSSFWMDKADISLIVPYI
GLANVGNTEAKGNFENAFETAGASILLEFPELLIPVYGAFLESEYIDNKNKIETI
NSALTDRKEDIMYGLLVAQLSTVQFYTIKEGMYKALNYQAQALEEIIKYKYNII
YSERKSNINIDFNVNSKLNENQAIIDNNINNECSVSLMKMKMPLAVELKILDF
DNTLRKLLNLTIDENKLGISAIEYKSKVDKLYKTSIPFDLSVTNNNTILIEIFNKY
NSDILLNLTILNLRVDNKLIDLGSYGARVEYDGVKLNKQKFKLTSSANSKIRVION
QNILFNSMELDFSVFWIRIPKYKNDGIONYIHNEYTLINCMKNSGKISIRGNMII
WTLIDINKISVFEYSIKEDISEYINRWFVTITNNSDKIYINKKLESHIDIRD
IREVIANDEIIFKLDGNIDRTQFIMWKYFISFNTLSQSNIEEIIKYQISSEYLDKDF
GNPLMYNKEYTFNAGNKNYSIKLKQDSPVGEILTRSKYQNSKYINRDLYIGSKFI
IRKSNQSQIINDDIVRKEDYIYLDFFNLNQEWRYIYKYFKKEEKLFLAPISDSDEF
YNTIQIKVDQPTVSCOLLFKKDEESTDEIGLIGHRFYESGIVFKEYKDYFCISKW
YLKWKRPYNSKGCNNQFIPKQBGWTE"
```

BASE COUNT 1611 a 370 c 615 g 1280 t
ORIGIN

```
alignment_scores:      Quality: 2231.00      Length: 438  
                        Ratio: 5.200      Gaps: 0  
Percent Similarity: 97.945      Percent Identity: 94.977
```

alignment_block:
US-09-910-186A-8 x AF295926 ..

Align seg 1/1 to: AF295926 from: 1 to: 3876

```
3 AsnLysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeuArgTy 19
|||||
2560 AATAAATAATACGATATTTAAATAATATATATCTTAATAATTAGATA 2609
|||||
19 rLysAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluV 36
|||||
2610 TAGGGATAATAAGTAAATAGATTTATCAGGATATGGGGCAAGGTAGAG 2659
|||||
36 alTyrAspGlyValGluLeuAsnAspLysAsnGlnPheLysLeuThrSer 52
|||||
2660 TATATGATGGGGTCAAGCTTAATGATAAAAAATCAATTTAAATTAACTAGT 2709
|||||
53 SerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePheAs 69
|||||
2710 TCAGCAATAAGTAGATAGATGATGATCAAAATCAGAAATATCATATTTAA 2759
|||||
69 nSerValPheLeuAspPheSerValSerPheTrpIleArgIleProLysT 86
|||||
2760 TAGTATGTCCTTCATTTTACCGTTAGCTTTTGGATAAAGAAATACCTAAAT 2809
|||||
86 yLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIleIle 102
|||||
2810 ATAAGAAATGATGTATACAAAAATATATTCATAATGAATATACAAATAAT 2859
|||||
103 AsnCysMetLysAsnAsnSerGlyTyrLysIleSerIleArgGlyAsnAr 119
|||||
2860 AATTGTGATAAAATATCTCGATGGAAAAATATCTATTAGGGGTAATAT 2909
|||||
119 gIleIleTrpThrLeuIleAspIleAsnGlyLysThrLysSerValPheP 136
|||||
2910 GATAAATAGGACCTTAATTCATATAAATGAAAAATCAAAATCAGTATTT 2959
|||||
136 heGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTrpPhe 152
|||||
2960 TTCAATATAGCATAAAGAAAGATATATCAGAGTATATAAATAGATGTTT 3009
|||||
```

seq_name: gb_ba:AF300465

```
153 PheValThrIleThrAsnAsnLeuAsnAsnAlaLysIleTyrIleAsnGI 169
|||||
3010 TTTGTAACCTATTACTAATAATTCGGATAACGCATAAAATTTATATTAATGG 3059
|||||
169 yLysLeuGluSerAsnThrAspIleLysAspIleArgGluValIleAlaA 186
|||||
3060 TAAGCTAGAATCATATAGATATATTAGAGATATAAGAGAAGTTATTGCTA 3109
|||||
186 snGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThrGlnPhe 202
|||||
3110 ATGATGAAATAAATAATTAAATTAGATGCTAATATAGATAGACACAAATTC 3159
|||||
203 IleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGlnSerAs 219
|||||
3160 ATTTGGATGAAATATTTCAGTATTTTAAATACGGAATTAAGTCAATCAA 3209
|||||
219 nIleGluGluArgTyrIlysIleGlnSerTyrSerGluTyrIleLysAsp 236
|||||
3210 TATTTGAAGAATAATATAAAATTTCAATCATATAGCGAATATTTAAAGATT 3259
|||||
236 heTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPheAsnAla 252
|||||
3260 TTTGGGGAAATCCTTTAATGTACATAAAGAAATATATATATGTTTAAATGCC 3309
|||||
253 GlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProValGlyGI 269
|||||
3310 GGGAATAAAAAATTCATATATTAACATAAGAAAGATTCACCTGTAGTGA 3359
|||||
269 uIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleAsnTyrA 286
|||||
3360 AATTTTAACACGTAGCAATAATATCAAAATTTCCAAATATATAAATATA 3409
|||||
286 rGAspLeuTyrIleGlyGluLysPheIleIleArgArgLysSerAsnSer 302
|||||
3410 GAGATTTATATATTGGAGAAAAATTTATTATAGAGAAAGTCAAAATCT 3459
|||||
303 GlnSerIleAsnAspIleValArgLysGluAspTyrIleTyrLeuAs 319
|||||
3460 CAATCATATAATGATGATAGTAGTAGAAGAAAGAGATTTATATATCTAGA 3509
|||||
319 pPhePheAsnLeuAsnGlnGluTrpArgValTyrThrTyrLysTyrPheL 336
|||||
3510 TTTTATTTTAAATTTAAATCAAGAGTGGAGAGTATATATCTATAAAATATT 3559
|||||
336 yLysGluGluGluLysLeuPheLeuAlaProIleSerAspSerAspGlu 352
|||||
3560 AGAAAGAGGAGAAAAATTTGTTTGTAGCTCTATAAGTGTATCTCGATGAG 3609
|||||
353 LeuTyrAsnThrIleGlnIleLysGluTyrAspGluGlnProThrTyrSe 369
|||||
3610 TTTTACAATACTATACAAATAAAGAAATATGATCAACAGCAACATATAG 3659
|||||
369 rCysGlnLeuLeuPheLysLysAspGluGluSerThrAspGluIleGlyL 386
|||||
3660 TTGTCAGTTGCTTTTAAAAAAGATGAAGAAAGTACTGATGAGATAGGAT 3709
|||||
386 euIleGlyIleHisArgPheTyrGluSerGlyIleValPheGluGluTyr 402
|||||
3710 TGATTTGTTATTCATCGTTTCTACGAATCTGGAATTTGATTAAAGAGTAT 3759
|||||
403 LysAspTyrPheCysIleSerLysTrpTyrLeuLysGluValLysArgLy 419
|||||
3760 AAAGATATTTTGTATAGTAAATGCTACTTAAAGAGGTAAAAAGGAA 3809
|||||
419 sProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheIleProLysAspG 436
|||||
3810 ACCATATAAATTCAAAATTTGGGATGTAATTTGCCAGTTTATTCTCTAAAGATG 3859
|||||
436 luGlyTrpThrGlu 440
|||||
3860 AAGGGTGGACTGAA 3873
```

```

seq_documentation_block:
LOCUS      AF300465          3876 bp    DNA        linear        BCT 08-Oct-2001
DEFINITION Clostridium botulinum isolate 588 type B cryptic neurotoxin gene,
            complete cds.
ACCESSION  AF300465
VERSION    AF300465.1  GI:15982936
KEYWORDS
SOURCE     Clostridium botulinum.
            Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
            Clostridium.
REFERENCE  1 (bases 1 to 3876)
            Kirma, N., Ferreira, J.L. and Baumstark, B.R.
            Characterization of six type A strains of Clostridium botulinum
            that contain type B toxin gene sequences
            Unpublished
REFERENCE  2 (bases 1 to 3876)
            Kirma, N., Ferreira, J.L. and Baumstark, B.R.
            Direct Submission
JOURNAL   Submitted (28-AUG-2000) Department of Biology, Georgia State
            University, P.O. Box 4010, Atlanta, GA 30302-4010, USA
ATTURES   location/Qualifiers
            1. .3876
            /organism="Clostridium botulinum"
            /isolate="588"
            /db_xref="taxon:1491"
            /country="USA: Ohio"
            /note="isolated from stool sample in 1976
            type: Ab"
            1. .3876
            /codon_start=1
            /transl_table=1
            /product="type B cryptic neurotoxin"
            /protein_id="AA11498.1"
            /db_xref="GI:15982937"
            /translation="MPVTINNFNYNDPIDNNIMPEPPFARGMGRYKAFKIDTRIW
            IIPFTFGYPKDFNKGSGIFNRDCEYDDPYLATNDKKNIFLOTMIKLPNRKISK
            PLGEKLEMTINGIPVGLDRRVLPEEFNTNIAVTVNKLISNPFGEVERKKGIFANLII
            FGPGVLNETIDIGIONHFASREGFGGMOKFCPEYVSFVNNQENKNGASTFNR
            GYSPDALILMHLIHLVGLGIVKYNLDLPVFNPKKFFMQSDATAQAEELYFGGQD
            PSIISSDTSKIDYKLVQNRGIVLRNKVLVCISDPNININIKKPKDKYKFVEDS
            EGYSDIVSFDEKLYKSLMPGFTETNIAENYKIKTRASFSDSLPVKKNLLDNEIY
            TIEGNSIDKMEKERYGQNKAINQAEIEISKEHLAVYKIQMCKSVKAPGICIDVD
            NEDLFTADKNSDDLSKNERIAYTONNYIENDESINELIOTLISKEIUPSNT
            ESITDENYVVPVKOPAKKIFTDENTIFQYLYSOTEPDLDIRDISLTSFDDALLFS
            NKYSPFSDMYIKTANKVVEAGLFGAGVKQIVNDFVIEANKSSTMBKDIADISLIVPI
            GLALNVGTAKGNFENAFEGASILLFEIPELLIPVVGAFLLSEYIDNKNKILIEI
            NSALTQRDEKIDMYGLIVAQWLSTVNTQFTYIKEGYKALNYQAQALEEIKRYKNI
            YSEKRSNINIDFVNSKLNQINQAINNFINECSVYLMKRMPIPLAEKLLDF
            DNTLRKNLLNYIDENKLYLGSAEYKSKVDYKLTSPFDLSTYNTNTLIEFNKY
            NSDILANILNLRNKLIDLSYGAKVEYDGVKINDKNOPKLTSSANSKTRVTQN
            ONIIFNSMFLDSVSWIRPKNGIQNYIHNEYTIINCMKNNSGWLKISIRGNMII
            WTLIDINGKTSVFFESYKEDISEYNIRWFFVTITNNSDNAKIYINGKLESHDIRD
            IREVIANDEITFLKLDGNDIRTOPIMWKYFISFNTLSQSNIEEYIQSYSEYKDFW
            GNPLMYNKYMFNAGNKNSYIKLKDSSVSGEILLTRSKYNSKSYINRYDLYIGERFI
            IRKNSQSNDIDIVKEDYIYLDENLNQEWRYMYKFKKEEKLFLAPISDSDEF
            YNTIQIKVEYDEQPTYSQCLLFKDEESTDEIGLIGTHRFVESGIVPEKDYFCISKW
            YLKEVRKPKNSKLGKNWGFIPKDEGWE"
BASE COUNT      1611 a      367 c      616 g      1282 t
ORIGIN
alignment_scores:
  Quality: 2229.00      Length: 438
  Ratio: 5.196          Gaps: 0
  Percent similarity: 97.945      Percent identity: 94.977
alignment_block:
US-09-910-186a-8 x AF300465
Align seg 1/1 to: AF300465 from: 1 to: 3876

```

```

3 AsnLysTyrAsnSerGluLeuAsnAsnIleLeuAsnLeuArgTyr 19
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2560 AATAATAATAGCGATATTTAAATAATATATCTTAAATTTAAGATA 2609
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
19 rLysAspAsnAsnLeuLeuAspLeuSerGlyTyrGlyAlaLysValGluV 36
I:::|||||:|||||:|||||:|||||:|||||:|||||:|||||
2610 TAGGGATAATAAGTTAATAGATTATATCAGGATATGGGCAAGAGTAGAGG 2659
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
36 alTyrAspGlyValGluLeuAsnAspLysAsnGlnPheIysLeuThrSer 52
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2660 TATATGATGGGTCAAGCTTAATGATAAAATCAATTTAAATTAACCTAGT 2709
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
53 SerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePheAs 69
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2710 TCAGCAATAGTAGAATAGAGTGAAGTCAAAATCAGATATCATATTTAA 2759
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
69 nSerValPheLeuAspPheSerValSerPheTrpIleArgIleProLysT 86
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2760 TAGTATGTTCTTGTATTTAGCGTTAGTTTGGGATAAGAAATACCTAAAT 2809
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
86 rLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIleIle 102
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2810 ATAAGAAATGATGATACAAAAATATATTCATPAATGAATATACAAATAAT 2859
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
103 AsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgGlyAsnAr 119
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2860 AATTGATGAAAATAAATTCGTGATGGAATAATCTATTAGGGGTAATAT 2909
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
119 gIleIleTrpThrLeuIleAspIleAsnGlyLysThrLysSerValPheP 136
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2910 GATAATATGGACTTTAATGTATATAAATGGAATAATCAAAATCAGTATTTT 2959
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
136 heGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTrpPhe 152
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
2960 TTGAATATAGCATAAAGAAAGATATATCAGAGTATATAAATAGATGGTTT 3009
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
153 PheValThrIleThrAsnAsnLeuAsnAsnAlaLysIleTyrIleAsnGl 169
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3010 TTTGTAACTATTACTAATAATTCGGATAACGCTAAATTTATATTAATGG 3059
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
169 yLysLeuGluSerAsnThrAspIleAspIleArgGluValIleAlaA 186
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3060 TAAAGTAGAATACACATATAGATATTAGAGATAAGAGAAGTATTGTCTA 3109
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
186 snGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThrGlnPhe 202
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3110 ATGATGAATAATAATTTAAATAGATGGTAAATATAGATAGACACAGTTC 3159
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
203 IleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGlnSerAs 219
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3160 ATTTGGATGAATAATTTTCAGTATTTTAAACGGAATTAAGTCAATCAAA 3209
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
219 nIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeuLysAspP 236
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3210 TATTGAAGAAATATATAAAATCAATCATATACGCAATATTTAAAAAGATT 3259
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
236 heTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPheAsnAla 252
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3260 TTTGGGAAATCCCTTAATGTACAAATAAGAAATATATATGTTTAAATGCC 3309
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
253 GlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProValGlyGl 269
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3310 GGAATAAAATTCATATATATAAACTAAAGAAAGATTTCATCTGTAGGTGA 3359
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
269 uIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleAsnTyrA 286
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3360 AATTTTAAACACGTAGCAAAATATAAATCAAAATTTCCAAATATATAAATATA 3409
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
286 rGAspLeuTyrIleGlyGluLysPheIleIleArgArgLysSerAsnSer 302
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3410 GAGATTTATATATTCGAGAAAATTTATATAGAAGAAAGTCAAAATTC 3459
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
303 GlnSerIleAsnAspIleValArgLysGluAspTyrIleTyrLeuAs 319

```

```
|||||
3460 CAATCTATAATGATGATATAGTTAGAAAAGAGATATATATATCTAGA 3509
|||||
319 pPheAsnLeuAsnGlnGluTrpArgValTyrThrTyrLysTyrPheL 336
|||||
3510 TTTTAAATTTAAATCAAGAGTGGAGAGATATATATATAATATTTA 3559
|||||
336 yLysGluGluGluLysLeuPheLeuAlaProIleSerAspSerAspGlu 352
|||||
3560 AGAAGAGAGAGAAAATTTGTTTGTAGCTCCTATAGTCTGATGAG 3609
|||||
353 LeuTyrAsnThrIleGlnIleLysGluTyrAspGluGlnProThrTyrse 369
|||||
3610 TTTTACAACTACTACAAAATAAAGAAATATGATGAACAGCAACATATAG 3659
|||||
369 rCysGlnLeuLeuPheLysLysAspGluGluSerThrAspGluIleGlyL 386
|||||
3660 TTGTGAGTTGCTTTTAAAAAAGATGAAGAAAGTACTGATGATAGGAT 3709
|||||
386 euIleGlyIleHisArgPheTyrGluSerGlyIleValPheGluGluTyr 402
|||||
3710 TGATTTGGTATTCATCGTTCTACGATTCGGAATGTTATTAAGAGTAT 3759
|||||
403 LysAspTyrPheCysIleSerLysTyrTyrLeuLysGluValLysArgLy 419
|||||
3760 AAAGATTATTTGTATAGTAATGTTACTTAAAGAGGTAAGAAAGAA 3809
|||||
419 sProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheIleProLysAspG 436
|||||
3810 ACCATATAATCAAAATGGGATGTAATGGCAGTTTATTCCTAAAGATG 3859
|||||
436 luGlyTrpThrGlu 440
|||||
3860 AAGGGTGGACTGAA 3873
```

seq_name: gb_ba:AF300466

```
seq_documentation_block:
LOCUS AF300466 3876 bp DNA linear BCT 08-OCT-2001
DEFINITION Clostridium botulinum isolate 593 type B cryptic neurotoxin gene,
complete cds.
```

```
ACCESSION AF300466
VERSION AF300466.1 GI:15982938
```

KEYWORDS

SOURCE

Clostridium botulinum.

Clostridium botulinum

Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;

Clostridium.

1 (bases 1 to 3876)

Kirma,N., Ferreira,J.L. and Baumstark,B.R.

Characterization of six type A strains of Clostridium botulinum

that contain type B toxin gene sequences

Unpublished

2 (bases 1 to 3876)

Kirma,N., Ferreira,J.L. and Baumstark,B.R.

Direct Submission

Submitted (28-AUG-2000) Department of Biology, Georgia State

University, P.O. Box 4010, Atlanta, GA 30302-4010, USA

Location/Qualifiers

1. .3876

/organism="Clostridium botulinum"

/isolate="593"

/db_xref="taxon:1491"

/country="USA: Georgia"

/note="isolated from dog feces in 1976

type: A(B)"

1. .3876

/codon_start=1

/transl_table=11

/product="type B cryptic neurotoxin"

/protein_id="AAL11499.1"

/db_xref="GI:15982938"

/translation="MPVTINNPNYNDPIDNNNIMPEPPFARGMGRYKAFKITDRIW

CDS

```
IIIPRYTFGYKPFNFSSGIFNRDVCVEYDYLNTNKKNIQLQTMIKLFRKIKK
PLGKLLLENIINGIPLGDRVPLEEFNTVVKLIINPCEVGRKKGFANLII
FGPGVLENETIDIGIONHPASREGGIMQKECEYVSFVNNVOENKASIFNRR
GYFSDPALILMHELHVLHGLYGLKYNLPIVNEKKEFFMSTDAIQAEELYTFGGQD
PSIIISPDSKSIYDKVLQNFQFVDRLNKLVLCISDPNININIKNPKDKYKFEVDS
EGYSKIDVESDKLYKSLMFGFTTETNAENYIKTRASYSDSLPPVKIKLLDNEYI
TIERGFNISDKNMEKEYRGONKAIKQAEYEISKEHLAVYKIQCKSVKAPGICIDVD
NEDLFTIADKNSFSDLLSKNERIAYNTQNNYIENDFINELIDTLTISKTELPSENT
ESLDFNVVYVPVYKKOPAKIKFTDENTIFOYLYSOTFPLDIRISLTSFDDALLES
NKYSFSPMDYIKTANKVVEAGLFGWVKQIVNDFVIEANKSSMDKIDSLISLVPII
GLALVNGNETAKGNFENAFIAGASILLEPILLIPVVGAFLESYIDNKKIETI
NSALTDRKEDWIDMYLIVAQWLSTVNTQFTYIKEGYKALNYQAQALEEIKYKINI
YSEKERSNINIDFNDVNSKLNENQAINNINNEINCEVSYLAKMKMPLAVEKLLDF
DNTLRKNLNYIDENKLYLGSAEYKSVKDYKLTSTIPDLSTVTNNITLIEIPNKY
NSDILNLIILNRYDNKLIDLSYGAKVENVYDKVNDKNQKLTSSANSKIRVTQN
QNTIFNSMFLDFSVPWIRIPKNDGQIONIHREYTIINCCKNSGWSKISIRGNMII
WTLIDINGKIKSVFEYSIKEDISEYINRWFFVTITNNSDAKIYINKLSEHIDIRD
IREVIANDEIFIKLDGNIDRTQFIWMKYFSIFNTELSQSNIEEYIKQSYEYLKDFW
GNPLMYKSYMFMAGNKNYSYIKLKDSVSGEILTRSKYNSQNSKYINRYDLYIQEKFI
IRKNSNSQINDDIVRKEDYIYLDFFNLNDEMRVYMYKYFKKEEKLFLAPISDSEF
YNTIQIKEYDEOPTYSCQLLEKDEESTDEIGLGIHRFYESGIVFKEYKDYFCISKW
YLKEVKKRPYNSKLGCMNQFIPKDEGWTE"
```

BASE COUNT 1611 a 367 c 616 g 1282 t
ORIGIN

alignment_scores:

Quality: 2229.00 Length: 438

Ratio: 5.196 Gaps: 0

Percent Similarity: 97.945 Percent Identity: 94.977

alignment_block:

US-09-910-186A-8 x AF300466

Align seg 1/1 to: AF300466 from: 1 to: 3876

```
3 AsnLysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeuArgTy 19
|||||
2560 AATAAATAATAGCGATATTTAAATAATATATATCTTAAATTAAGATA 2609
|||||
19 rLysAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluV 36
|||||
2610 TAGGGATAATAAGTTAATAGATTTATCAGGATATGGGCAAGGTAGAGG 2659
|||||
36 aLTYrAspGlyValGluLeuAsnAspLysAsnGlnPheLysLeuThrSer 52
|||||
2660 TATATGATGGGGTCAAGCTAATGATAAAAATCAATTTAAATTAAGTAGT 2709
|||||
53 SerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePheAs 69
|||||
2710 TCAGCAAAATAGTAAGATTAGAGTGACTCAAAATCAGAATATCATATTTAA 2759
|||||
69 nSerValPheLeuAspPheSerValSerPheTrpIleArgIleProLysT 86
|||||
2760 TAGTATGTTCCCTTGATTTAGCGTTAGTTTGGATAAAGAAATACCTAAAT 2809
|||||
86 yLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIleIle 102
|||||
2810 ATAAGAATATGTTATACAAAATATATATCAATAATGAATACATAAATAAT 2859
|||||
103 AsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgGlyAsnAr 119
|||||
2860 AATTGTATCAAAAATAATCTGGATGAAAATATCTATTAGGGGTAATAT 2909
|||||
119 gIleIleTrpThrLeuIleAspIleAsnGlyLysThrLysSerValPheP 136
|||||
2910 GATAATATGGACTTTAATTCATATAAATGAAAAATCAAAATCAGTATTTT 2959
|||||
136 heGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTrpPhe 152
|||||
2960 TTGAATATAGCATAAAAGAGATATATACAGATATATAAATAGATGGTTT 3009
|||||
153 PheValThrIleThrAsnAsnLeuAsnAsnAlaLysIleTyrIleAsnGln 169
```

```
|||||.....:::|||||.....
3010 TTTGTAACATATTAATAATTCGATACGCTAAATTTATATTAATG 3059
169 YLLeuGluSerAsnThrAspIleLeuAspIleArgGluValIleAla 186
|||||.....:::|||||.....
3060 TAAGCTAGAAACACATATAGATATTAGATATAGAGAAGTATTGCTA 3109
186 snGlyGluIleLeuPheLysLeuAspGlyAspIleAspArgThrGlnPhe 202
|| .....:::|||||.....
3110 ATGATGAATAATATTTAAATAGATGCTATATAGATAGAACACAGTTC 3159
203 IleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGlnSerAs 219
|||||.....:::|||||.....
3160 ATTTGGATGAATATTTTCAGTATTTTAAATACGGAATTAAGTCAATCAA 3209
219 nIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeuLysAsp 236
|||||.....:::|||||.....
3210 TATTGAAGAATATATAAAATTCATCATATAGCGAATATTTAAAGATT 3259
236 heTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPheAsnAla 252
|||||.....:::|||||.....
3260 TTTGGGAATCCITTAATGACATAAAGAAATATATATGTTAATCGG 3309
253 GlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProValGlyG 269
|||||.....:::|||||.....
3310 GGGATAAAATTCATATATTAACATAAAGAAAGATTTCATCTAGGTGA 3359
269 uIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleAsnTyrA 286
|||||.....:::|||||.....
3360 AATTTTAACACGTAGCAATATAATCAAAATTCCAATATATAAATATA 3409
286 rGAspLeuTyrIleGlyGluLysPheIleIleArgArgLysSerAsnSer 302
|||||.....:::|||||.....
3410 GAGATTTATATATTGGAGAAAAATTTATATAGAGAAAGTCAATCTCT 3459
303 GlnSerIleAsnAspAspIleValArgLysGluAspTyrIleTyrLeuAs 319
|||||.....:::|||||.....
3460 CAATCATATAATGATGATATAGTTAGAAAGAAAGATTATATATCTAGA 3509
319 pPhePheAsnLeuAsnGlnGluTrpArgValTyrThrTyrLysTyrPheL 336
|||||.....:::|||||.....
3510 TTTTAAATTTAAATCAATCAAGAGTGGAGAGTATATATGATATAAATTTA 3559
336 ysLysGluGluGluLysLeuPheLeuAlaProIleSerAspSerAspGlu 352
|||||.....:::|||||.....
3560 AGAAGAGGAGAGAAAAATGTTTGTAGCTCTATTAAGTATGATCTCATG 3609
353 LeuTyrAsnThrIleGlnIleLysGluTyrAspGluGlnProThrTyrSe 369
:::|||||.....:::|||||.....
3610 TTTTACAATACTATACAAATAAAAGAAATATGATGAACAGCCAAATATAG 3659
369 rCysGlnLeuLeuPheLysLysAspGluGluSerThrAspGluIleGlyL 386
|||||.....:::|||||.....
3660 TTGTCAGTTGCTTTTAAAAAGATGAAGAAAGTACTGATGAGATAGGAT 3709
386 euIleGlyIleHisArgPheTyrGluSerGlyIleValPheGluGluTyr 402
|||||.....:::|||||.....
3710 TGATTTGATTCATCGTTCTACGAATCTGGAATGTTATTAAGAGTAT 3759
403 LysAspTyrPheCysIleSerLysTyrTrpTyrLeuLysGluValLysArgLy 419
|||||.....:::|||||.....
3760 AAAGATTATTTTGTATTAAGTAAATGGTACTTAAAGAGGTAAGAAAGGAA 3809
419 sProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheIleProLysaspG 436
|||||.....:::|||||.....
3810 ACCATATAATTAATCAAAATTTGGATGTAATGGCAGTTTATTCCTCAAGATG 3859
436 luGlyTrpThrGlu 440
|||||.....:::|||||.....
3860 AAGGTGGACTGAA 3873
seq_name: gb_ba:CBY13630
```

```
seq_documentation_block:
LOCUS CBY13630 11170 bp DNA linear BCT 25-OCT-1998
DEFINITION Clostridium botulinum HA-70, HA-17, HA-33, P-21, ntnh, bont genes.
ACCESSION Y13630
VERSION Y13630.1 GI:3805779
KEYWORDS bont gene; HA-17 gene; HA-33 gene; HA-70 gene; ntnh gene; P-21 gene.
SOURCE Clostridium botulinum.
ORGANISM Clostridium botulinum
Bacteria: Firmicutes; Bacillus/Clostridium group; Clostridiaceae; Clostridium.
REFERENCE 1 (bases 1 to 11170)
AUTHORS Santos-Buelga, J.A., Collins, M.D. and East, A.K.
TITLE Characterization of the genes encoding the botulinum neurotoxin complex in a strain of Clostridium botulinum producing type B and F neurotoxins
JOURNAL Curr. Microbiol. 37 (5), 312-318 (1998)
MEDLINE 98440323
REFERENCE 2 (bases 1 to 11170)
AUTHORS Santos-Buelga, J.A.
TITLE Direct Submission
JOURNAL Submitted (05-JUN-1997) J. A. Santos-Buelga, BBSRC Institute of Food Research, Earley Gate, Whiteknights Road, Reading, RG6 6BZ, UK
FEATURES
source
1..11170
/organism="Clostridium botulinum"
/strain="CDC 3281 (ATCC 43757)"
/db_xref="taxon:1491"
/clone="PAET98-99"
/clone="pJSB1-7"
/clone="pJSB20"
/complement(1..1326)
/gene="HA-70"
/complement(1340..1780)
/gene="HA-17"
/complement(1340..1780)
/gene="HA-17"
/codon_start=1
/transl_table=11
/protein_id="CAA73963.1"
/db_xref="GI:3805780"
/translation="MNSSIKKIYNDIOEKVINVSDDIDLAGNVYVRGSGWILSRQN
QILGSGVINSSTGIVGLRVNDNAIPIYPTSPFNEEYIKNNIQVFNFTBANQIP
IGVEFKTAPSNKNLYQIYVRIEIKVLQHEIIEIRAVLYPSLGYVTKSEFNP
EKINKDFYFLTNDKILNQFYKILETKNPTNNIFNSKVSSTQRLVPSNGLYV
INKDCYIRTNDDLIGLLIEAGSSGSIIOPLRNTTLPFTTSNDTKFSQOYTEER
LKDAFNQLENTSTSLKPEEAPSKNICIKAYTEKVELIDYQNGSVLNKAEYVL
PSLGCYEVNAPSPSESVKMQVAEDGFQNGPEEIVGVIDPDSERIOBINTAIDSN
YIYSPGIVNNNPFIILFTVNTTGIYKINAQNNLPLPKIYEATGSGNRNN"
gene
CDS
complement(1340..1780)
/gene="HA-17"
/complement(1340..1780)
/gene="HA-17"
/codon_start=1
/transl_table=11
/protein_id="CAA73964.1"
/db_xref="GI:3805781"
/translation="MSVERTFLPNGVNNTKISFSGSLYNPVSLSLTFSSNESSANNOK
WVVEYMAENRCFKISNVAEPNKLYSDNFCFISLDSLSNRCWFKPIAVNTYIMLSL
NKVNEIDYAWDIYDTNENILSQPLLELPNFDIYNSQMFKLEK1"
gene
CDS
complement(1843..2724)
/gene="HA-33"
/complement(1843..2724)
/gene="HA-33"
/codon_start=1
/transl_table=11
/protein_id="CAA73965.1"
/db_xref="GI:3805782"
/translation="MEHYSVIONSLNDKIVTISCRA DTNLFFYQVAGNVSLSFOQTRNY
LERWRIIYDSNKAAYKIKSDIHTNLVLTNAPTNTISTQDSNADNOYWLKLDIG
NSNFIATSYKNPNLYADTVARNKLSTLNNSNYIKFIEDYIISDENNTCKIKSPI
LDNRKVQOQVATNTNLNVLTYWDYGRNQKWTIRYNEEKAAYQFFNTILSGVLTWIFS
NGNTVRVSSSDONNDNQWLINPVSDDTDEYTTITNLRDITTKALDLYNSQTANGTAIQ
VFNSSNGDGNQKNIRNP"
```


253 GlyAsnLysAsnSerTyrIleLeuLysLeuLysLysAspSerProValGlyGI 269
 10577 GCGAATAAAATTCATATATTAACTAAAGAAAGATTTCATCTGAGGTCA 10626
 269 urleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleAsnTyrA 286
 10627 AATTTTACACGTACCAATATAATCAAAATTCCAAATATATAAAATATA 10676
 286 rGAspLeuTyrIleGlyGluLysPheIleIleArgArgLysSerAsnSer 302
 10677 GAGATTATATATTGGAGAAAATTTATATATAAGAAAGAGCTCAAAATCT 10726
 303 GlnSerIleAsnAspIleValArgLysGluAspTyrIleTyrLeuAs 319
 10727 CAATCTATAAATGATGATATAGTTAGAAAGAAAGATTATATATCTAGA 10776
 319 pPhePheAsnLeuAsnGlnGluTyrArgValTyrThrTyrLysTyrPheL 336
 10777 TTTTTTTTAAATCAAGAGTGGAGAGTATATATGATATAAATATTTTA 10826
 336 ysLysGluGluGluLysLeuPheLeuAlaProIleSerAspSerAspGlu 352
 10827 AGAAAGAGAGAAANAATGTTTGTAGTCTCTATAAGTGAATTCGATCAG 10876
 353 LeuTyrAsnThrIleGlnIleLysGluTyrAspGluGlnProThrTyrSe 369
 10877 TTTTCACTACTATACAAATAAAAGAAATATGATGAACAGCCCAACATATAG 10926
 369 rCysGlnLeuLeuPheLysLysAspGluSerThrAspGluIleGlyL 386
 10927 TTGTCAGTTGCTTTTAAAGAAAGATGAAGAAAGTACTGATGAGATAGGAT 10976
 386 euIleGlyIleHisArgPheTyrGluSerGlyIleValPheGluGluTyr 402
 10977 TCAATGGTATTCATCGTTCTTACGNAATCGNAATGATTTAAAGAGTAT 11026
 403 LysAspTyrPheCysIleSerLysTrpTyrLeuLysGluValLysArgLy 419
 11027 AAAGATTATTTTGTATAGTAAATGGTACTTAAAGAGGTAAAGAGAA 11076
 419 sProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheIleProLysAspG 436
 11077 ACCATATAATCAAAATTTGGATGATTAATGGCAGTTTATCTCTAAAGATG 11126
 436 luGlyTrpThrGlu 440
 11127 AAGGTGGACTGAA 11140

seq_name: gb_ba:AF300467

seq_documentation_block:

CUS AF300467 3869 bp DNA linear BCT 08-OCT-2001
 DEFINITION Clostridium botulinum isolate 519 type B cryptic neurotoxin-like
 gene, complete sequence.

ACCESSION AF300467

VERSION AF300467.1 GI:15982940

KEYWORDS Clostridium botulinum.

SOURCE Clostridium botulinum

ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
 Clostridium.

REFERENCE

1 (bases 1 to 3869)

Kirma,N., Ferreira,J.L. and Baumstark,B.R.

Characterization of six type A strains of Clostridium botulinum

that contain type B toxin gene sequences

Unpublished

2 (bases 1 to 3869)

Kirma,N., Ferreira,J.L. and Baumstark,B.R.

Direct Submission

Submitted (28-AUG-2000) Department of Biology, Georgia State

University, P.O. Box 4010, Atlanta, GA 30302-4010, USA

NCBI staff are still waiting for submitters to provide appropriate

coding region information.

FEATURES
 source
 Location/Qualifiers
 1. 3869
 /organism="Clostridium botulinum"
 /isolate="519"
 /db_xref="taxon:1491"
 /country="USA: Alaska"
 /note="isolated from stool sample in 1976
 type: A(B)"
 misc_feature
 1. 3869
 /note="similar to type B cryptic neurotoxin"
 BASE COUNT 1607 a 365 c 614 g 1283 t
 ORIGIN

alignment_scores:
 Quality: 2210.00 Length: 438
 Ratio: 5.140 Gaps: 0
 Percent Similarity: 98.174 Percent Identity: 94.521
 alignment_block:
 US-09-910-186A-8 x AF300467 ..
 Align seg 1/1 to: AF300467 from: 1 to: 3869

3 AsnLysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeuArgTy 19
 2554 AATAAATAATAGCGATATTTTAAATAATATATATCTTTAAATTTAAGATA 2603
 19 rLysAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluV 36
 2604 TAGGGATAATAAGTTAATAGATTTTATCAGGATATGGGCAAGGTAGAGG 2653
 36 alTyrAspGlyValGluLeuAsnAspLysAsnGlnPheLysLeuThrSer 52
 2654 TATATGATGGGGTCAAGCTTAATGATAAAATCAATTTAAATTAAGTAGT 2703
 53 SerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePheAs 69
 2704 TCAGCAAAATAGTAGATAGAGTGAAGTCACTCAAAATCAGAAATATCATATTTAA 2753
 69 nSerValPheLeuAspPheSerValSerPheTrpIleArgIleProLysT 86
 2754 TAGTATGTTCTTGTATTTAGCGTTAGTTTTTGGATAAGAAATACCTAAAT 2803
 86 yrLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIleIle 102
 2804 ATAAGAATGATGGTATACAAAATATATTTATCATAATGAATATACATAATT 2853
 103 AsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgGlyAsnAr 119
 2854 AATTGTATGAAAAATAATCTCGGATGGAAATA.TCTATTAGGGGTAAATAT 2902
 119 gIleIleTrpThrLeuIleAspIleAsnGlyLysThrLysSerValPheP 136
 2903 GATATATGGACTTTAATGATATAAATGGAAAAATCAAAATCAGATTTTT 2952
 136 heGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTrpPhe 152
 2953 TTGAATATAGCATAAAACAAGATATATCAGAGTATATAAATAGATGGTTT 3002
 153 PheValThrIleThrAsnAsnLeuAsnAlaLysIleTyrIleAsnGl 169
 3003 TTTGTAACTATTACTATAATTCGATACCGCTAAAATTTATATTAATGG 3052
 169 yLysLeuGluSerAsnThrAspIleLysAspIleArgGluValIleAlaA 186
 3053 TAAGCTAGAAATCACATATAGATATTAGAGATATAAGACAGATTATTGCTA 3102
 186 snGlyGluIlePheLysLeuAspGlyAspIleAspArgThrGlnPhe 202
 3103 ATGATGAAATAATATTTAAATTAGATGCTAATATAGATAGAACACAATTC 3152
 203 IleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGlnSerAs 219


```
|||||
3153 ATTGGGATGAATATTTCAGCTGTTTTTAATACGGAATTAAGTCAATCAA 3202
|||||
219 nileGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeuLysAsp 236
|||||
3203 TATTGAAGAAAGATATAAAATTCATCATATAGCGAATATTTAAAGATT 3252
|||||
236 heTPrGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPheAsnAla 252
|||||
3253 TTTGGGGAATCCTTAAATGACAAATAAGAAATATATATGTTAAATGCG 3302
|||||
253 GlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProValGlyG 269
|||||
3303 GCGAATAAAATTCATATATTAACATAAGAAAGATTCCACCTGAGGTGA 3352
|||||
269 uileLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleAsnTyrA 286
|||||
3353 AATTTTAACACGAGCAATATATATCAAAATTCCAAATATATATAATATA 3402
|||||
286 rGAspLeuTyrIleGlyGluLysPheIleArgArgLysSerAsnSer 302
|||||
3403 GAGATTTATATATGGAGAAAATTTATATATAGAGAAAGTCAAAATCT 3452
|||||
303 GlnSerIleAsnAspIleValArgLysGluAspTyrIleTyrLeuAs 319
|||||
3453 CAATCTATAAATGATGATATAGTTAGAAAAGAGATTATATATATCTAGA 3502
|||||
319 pPhePheAsnLeuAsnGlnGluTyrPargValTyrThrTyrLysTyrPhe 336
|||||
3503 TTTTTTAAATTTAAATCAAGAGTGGAGATATATATCTATAAAATATTTA 3552
|||||
336 yLysGluGluGluLysLeuPheLeuAlaProIleSerAspSerAspGlu 352
|||||
3553 AGAAGAGGAGAGAAAATGTTTTAGCTCTATATAGTGATTTCTGATGAT 3602
|||||
353 LeuTyrAsnThrIleGlnIleLysGluTyrAspGluGlnProThrTyrSe 369
|||||
3603 TTTTACAATACTATACAAATAAAGAAATATGATGAACAGCAACATATAG 3652
|||||
369 rCysGlnLeuLeuPheLysLysAspGluGluSerThrAspGluIleGlyL 386
|||||
3653 TTGTCAGATTCCTTTTAAAGAAATGAAGAAAGTACTGATGATAGGAT 3702
|||||
386 eulleGlyIleHisArgPheTyrGluSerGlyIleValPheGluGluTyr 402
|||||
3703 TGATTTGGTATCATCGTTCTACGAATCTGGAATGGAATTTAAAGAGTAT 3752
|||||
403 LysAspTyrPheCysIleSerLysTyrTyrTyrLeuLysGluValLysArg 419
|||||
3753 AAAGATTATTTTGTATAAGTAAATGCTACTTAAAGAGGTAAAGAGAA 3802
|||||
419 sProTyrAsnLeuLysLeuGlyCysAsnThrGlnPheIleProLysAspG 436
|||||
3803 ACCATATAATTCAAAATTTGGGATGTAATTGGAAGTTTATTCCTAAAGATG 3852
|||||
436 luGlyTyrThrGlu 440
|||||
3853 AAGGGTGGACTGAA 3866

seq_name: gb_ba:AF300468

seq_documentation_block:
LOCUS AF300468
DEFINITION Clostridium botulinum isolate 667 type B cryptic neurotoxin-like
gene, complete sequence.
ACCESSION AF300468
VERSION AF300468.1 GI:15982941
KEYWORDS
SOURCE Clostridium botulinum.
ORGANISM Clostridium botulinum
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
Clostridium.
REFERENCE 1 (bases 1 to 3869)
```

```
AUTHORS Kirma,N., Ferreira,J.L. and Baumstark,B.R.
TITLE Characterization of six type A strains of Clostridium botulinum
that contain type B toxin gene sequences
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 3869)
AUTHORS Kirma,N., Ferreira,J.L. and Baumstark,B.R.
TITLE Direct Submission
JOURNAL Submitted (28-AUG-2000) Department of Biology, Georgia State
University, P.O. Box 4010, Atlanta, GA 30302-4010, USA
COMMENT NCBI staff are still waiting for submitters to provide appropriate
coding region information.
FEATURES
source
1. 3869
/organism="Clostridium botulinum"
/isolate="667"
/db_xref="taxon:1491"
/country="USA: Wisconsin"
/note="Isolated from stool sample in 1976
type: A(B)"
misc_feature
1. 3869
/note="similar to type B cryptic neurotoxin"
BASE COUNT 1607 a 365 c 614 g 1283 t
ORIGIN
alignment_scores:
Quality: 2210.00 Length: 438
Ratio: 5.140 Gaps: 0
Percent Similarity: 98.174 Percent Identity: 94.521
alignment_block:
US-09-910-186A-8 x AF300468 ..
Align seg 1/1 to: AF300468 from: 1 to: 3869
3 AsnLysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeuArgTy 19
|||||
2554 AATAAATATAATAGCGATATTTAAATAATATATCTTAAATTTAAGATA 2603
|||||
19 rLysAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluV 36
|||||
2604 TAGGGATAATAAGCTTAATAGATTTATCAGGATATGGGCAAGGTAGAGG 2653
|||||
36 alTyrAspGlyValGluLeuAsnAspLysAsnGlnPheLysLeuThrSer 52
|||||
2654 TATATAGGGGTCAAGCTTAATGATAAAATCAATTTAAATTAACTAGT 2703
|||||
53 SerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePheAs 69
|||||
2704 TCAGCAATAGTAAGATTACAGTGACTCAAAATCAGAAATATCATATTTAA 2753
|||||
69 nSerValPheLeuAspPheSerValSerPheTyrPheIleArgIleProLys 86
|||||
2754 TAGTATGTTCTCTGATTTTAGCGTTAGTTTGGGATAAGAAATACCTAAAT 2803
|||||
86 yLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIleIle 102
|||||
2804 ATAAGATGATGGTATACAAAATTTATATTCATATGAATATACAAATATT 2853
|||||
103 AsnCysMetLysAsnAsnSerGlyTyrLysIleSerIleArgGlyAsnAr 119
|||||
2854 AATTGTATGAAAATAATTTCTGGATGGAATA.TCTATTAGGGGTAATAT 2902
|||||
119 gIleIleTyrThrLeuIleAspIleAsnGlyLysThrLysSerValPheP 136
|||||
2903 GATAATATGGACTTTAATTCATATATAATGCAAAATCAAAATCAGTATTTT 2952
|||||
136 heGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTyrPhe 152
|||||
2953 TTGAATATAGCATAAAGAGATATATCAGAGTATATAAATAGATGGTTT 3002
|||||
153 PheValThrIleThrAsnAsnLeuAsnAsnAlaLysIleTyrIleAsnGl 169
|||||
```

```
3003 TTTGTAACCTATTACTAAATAAATTCGGATACGCTAAATAATTTATATTAATGG 3052
169 yLysLeuGluSerAsnThrAspIleLysAspIleArgGluValIleAlaA 186
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3053 TAAGCTAGATACATATAGATATTAGAGATATAGAGAGAGTATTGCTA 3102
186 sNGlyGluIlePheLysLeuAspGlyAspIleAspArgThrGlnPhe 202
|||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3103 ATGATGAATAATATTAAATTAGATGGTAATATAGATAGAACACAATTC 3152
203 IleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGlnSerAs 219
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3153 ATTTGATGAATAATTTTCAGTCTTTTATACGGAAATTAAGCAATCAAAA 3202
219 nIleGluIleArgTyrLysIleGlnSerTyrSerGluTyrLeuLysAspP 236
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3203 TATTGAAGAAAGATATAAAATTCATCATATATAGCGAATATTTAAAGATT 3252
236 heTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPheAsnAla 252
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3253 TTTGGGGAATCCTTTAATGTACATAAAGATAATATATATGTTAATGCG 3302
253 GlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProValGlyGl 269
3303 GGGATAAAATTCATATATTAACATAAAGAAAGATTCCACCTGTAGGTGA 3352
269 uIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleAsnTyrA 286
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
3353 AATTTTACACGTCAGCAATATATCAAAATTCCAAATATATATAATATA 3402
286 rGAspLeuTyrIleGlyGluLysPheIleIleArgArgLysSerAsnSer 302
3403 GAGATTTATATATTCGAGAAAAATTTATATAAGAAAGAAAGTCAAAATCT 3452
303 GlnSerIleAsnAspIleValArgLysGluAspTyrIleTyrLeuAs 319
3453 CAATCTATAATAGATATAGTATTAGAAAGAAAGATATATATATATCTAGA 3502
319 pPhePheAsnLeuAsnGlnGluTrpArgValTyrThrTyrLysTyrPheL 336
3503 TTTTAAAAATTAATCAAGAGTGGAGAGTATATATCTATAAATATTTTA 3552
336 ySLysGluGluLysLeuPheLeuAlaProIleSerAspSerAspGlu 352
3553 AGAAAAGAGAGAAAAATTTGTTTATAGCTCCTATAAGTGATTTCTGATGAT 3602
353 LeuTyrAsnThrIleGlnIleLysGluTyrAspGluGlnProThrTyrSe 369
3603 TTTTACATATCTATACAAATAAAGAAATATGATGACAGCCACATATAG 3652
369 rCysGlnLeuLeuPheLysLysAspGluSerThrAspGluIleGlyL 386
3653 TTGTCAATGCTCTTTTAAAAAAGATGAAGAAAGTACTGATGAGATAGGAT 3702
386 euIleGlyIleHisArgPheTyrGluSerGlyIleValPheGluGluTyr 402
3703 TGATTGGTATTCATCGTTCTACGAATCTGGAATTTGATTTTAAAGAGTAT 3752
403 LysAspTyrPheCysIleSerLysTrpTyrLeuLysGluValLysArgLy 419
3753 AAAGATATTTTGTATAGTAATAGTGTACTTTAAAGAGGTAAGAAAGNA 3802
419 sProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheIleProLysAspG 436
3803 ACCATATATTTCAAAATTTGGGATGTAATTTGGAAGTATTATTCCTAAAGATG 3852
436 luGlyTrpThrGlu 440
3853 AAGGGTGGACTGAA 3866
seq_name: gb_ba:AF300469
seq_documentation_block:
```

```
AF300469 3869 bp DNA linear BCT 08-OCT-2001
Clostridium botulinum isolate 13280 type B cryptic neurotoxin-like
gene, complete sequence.
ACCESSION AF300469
VERSION AF300469.1 GI:15982942
KEYWORDS Clostridium botulinum.
SOURCE Clostridium botulinum
ORGANISM Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
Clostridium.
REFERENCE 1 (bases 1 to 3869)
AUTHORS Kirma,N., Ferreira,J.L. and Baumstark,B.R.
TITLE Characterization of six type A strains of Clostridium botulinum
that contain type B toxin gene sequences
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 3869)
TITLE Direct Submission
JOURNAL Kirma,N., Ferreira,J.L. and Baumstark,B.R.
AUTHORS Submitted (28-AUG-2000) Department of Biology, Georgia State
University, P.O. Box 4010, Atlanta, GA 30302-4010, USA
COMMENT NCBI staff are still waiting for submitters to provide appropriate
coding region information.
FEATURES
source Location/Qualifiers
1..3869
/organism="Clostridium botulinum"
/isolate="13280"
/db_xref="taxon:1491"
/country="USA: Colorado"
/notes="isolated from peppers in 1972
type: A(B)"
misc_feature 1..3869
/notes="similar to type B cryptic neurotoxin"
BASE COUNT 1607 a 365 c 614 g 1283 t
ORIGIN
alignment_scores:
Quality: 2210.00 Length: 438
Ratio: 5.140 Gaps: 0
Percent Similarity: 98.174 Percent Identity: 94.521
alignment_block:
US-09-910-186A-8 x AF300469 ..
Align seg 1/1 to: AF300469 from: 1 to: 3869
3 AsnLysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeuArgTyr 19
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2554 AATAAATATATAGCGATATTTTAAATAATATATATCTTAAATTTAAGATA 2603
19 rLysAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluV 36
|:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2604 TAGGGATAATAAGTTAATAGATTTATCAGGATATGGGCAAGGTAGAGG 2653
36 alTyrAspGlyValGluLeuAsnAspLysAsnGlnPheLysLeuThrSer 52
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2654 TATATGATGGGTCAAGCTTAATGATAAAATCAATTTAAATTAACACTAGT 2703
53 SerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePheAs 69
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2704 TCAGCAATAGTAAAGATTAGAGTGACTCAAAATCAGATATCATATTTAA 2753
69 nSerValPheLeuAspPheSerValSerPheTrpIleArgIleProLysF 86
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2754 TAGTATGTTCTTGTATTTTAGCGTTAGTTTTTGGATAAGAAATACCTAAAT 2803
86 yLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIleIle 102
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2804 ATAAGAATGATGGTATACAAAATTTATATCAATAATGAATATACAATAAT 2853
103 AsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgGlyAsnAr 119
|||||.....:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
2854 AATTGTATGAAAAATAATTCTGGATGGAAATA.TCTATTAGGGGTAAATAT 2902
```

```
119 gillelletpThrLeuileAspIleAsnGlyLysThrLysSerValPheP 136
|||||
2903 GATAATATGGACITTAATTGATATAAATGGAATAATCAAAATCAGTATTTT 2952
136 heGluTyraAsnIleAArgLysAspIleSerGluTyriLeAsnAArgTrrPhe 152
|||||
2953 TTGAATATAGCATAAAAGAGATATATACAGATGATATAAATAGATGTTT 3002
153 PheValThrIleThrAsnAsnLeuAsnAlaLysIleTyriLeAsnGI 169
|||||
3003 TTGTAACTATTACTAATAATTCGGATAACGCTAAAATTTATATATATG 3052
169 yLysLeuGluSerAsnThrAspIleLysAspIleAArgLysValIleAla 186
|||||
3053 TAAGCTAGATACATATAGATATATAGATATAAGAGAGATTTATGCTA 3102
186 snGlyGluIlellePheLysLeuAspGlyAspIleAspAArgThrGlnPhe 202
|||||
3103 ATGATGAATAATATTTAAATTTAGATGTTTAAATATAGATAGACACAATTC 3152
203 IleTrrMetLysTyrrPheSerIlePheAsnThrGluLeuSerGlnSerAs 219
|||||
3153 ATTTGGATGAATAATTTCACTGTTTAAATACGGAATTAAGTCAATCAAA 3202
219 nileGluGluAArgTyrrLysIleGlnSerTyrrSerGluTyrrLeuLysAsp 236
|||||
3203 TATGAGAAGAAGATATAAAATTTCAATCATATAGCGAATATTTAAAGATT 3252
236 heTrrpGlyAsnProLeuMetTyrrAsnLysGluTyrrMetPheAsnAla 252
|||||
3253 TTTGGGAAATCCTTTAATCTACAATAAAGAATATATATGTTTAATGCG 3302
253 GlyAsnLysAsnSerTyrrIleLysLeuLysLysAspSerProValGlyGI 269
|||||
3303 GGGAAATAAAATTCATATATTAACATAAAGAAAGATTCACTGTAGGTGA 3352
269 uileLeuThrAArgSerLysTyrrAsnGlnAsnSerLysTyrrIleAsnTyrr 286
|||||
3353 AATTTAAACAGTAGCAATAATTAATCAAAATTCAAATATATAAATYATA 3402
286 rGAspLeuTyrrIleGlyGluLysPheIleIleAArgLysSerAsnSer 302
|||||
3403 GAGATTTATATATTTGGAGAAAAATTTATTAAGAGAAAGTCAAAATCT 3452
303 GlnSerIleAsnAspIleValAArgLysGluAspTyrrIleTyrrLeuAs 319
|||||
3453 CAATCTATAAATGATGATATAGTTAGAAAAAGAGATTATATATATCTAGA 3502
319 pPheAsnLeuAsnGlnGluTrrAArgValTyrrThrTyrrLysTyrrPheL 336
|||||
3503 TTTTAAAAATTAATCAAGAGTGGAGAGTATATATCTATAAATATTTTA 3552
336 yLysGluGluGluLysLysPheLeuAlaProIleSerAspSerAspGlu 352
|||||
3553 AGAAGAGGAAGAAAAATCTTTTAGCTCCTCTAAGTGAATTCGTGATGAT 3602
353 LeuTyrrAsnThrIleGlnIleLysGluTyrrAspGluGlnProThrTyrrSe 369
|||||
3603 TTTTACAATACTATACAAATTAAGAAATATGATGAACAGCAACATATAG 3652
369 rCysGlnLeuLeuPheLysLysAspGluGluSerThrAspGluIleGlyL 386
|||||
3653 TTGTCACTGCTCTTTTAAAAAGATGAAGAAAGTACTGATGAGATAGGAT 3702
386 eulleGlyIleHisAArgPheTyrrGluSerGlyIleValPheGluGluTyrr 402
|||||
3703 TGATTTGGTATTCATCGTTTCTACGAATCTGGAATTTGTAATTAAGAGTAT 3752
403 LysAspTyrrPheCysIleSerLysTyrrTyrrLeuLysGluValLysArgLy 419
|||||
3753 AAAGATTTATTTTGTATAAGTAATGCTACTTAAAAAGAGGTAAGAAAGAA 3802
```

```
419 sProTyraAsnLeuLysLeuGlyCysAsnTrrpGlnPheIleProLysAspG 436
|||||
3803 ACCATATAATTCAAATTTGGATGTAATTTGGAAGTTTATTCCTTAAGATG 3852
436 luGlyTrrpThrGlu 440
|||||
3853 AAGGGTGGACTGAA 3866
seq_name: gb_ba.CBBONTB
seq_documentation_block:
LOCUS CBBONTB 4051 bp DNA linear BCT 20-APR-1994
DEFINITION C.botulinum type B gene for neurotoxin.
ACCESSION X71343
VERSION X71343.1 GI:296148
KEYWORDS botNT/B gene; botulinum neurotoxin type B; neurotoxin type B.
SOURCE Clostridium botulinum.
ORGANISM Clostridium botulinum.
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
Clostridium.
REFERENCE
1 (bases 1 to 4051)
Hutson,R.A., Collins,M.D., East,A.K. and Thompson,D.E.
Nucleotide sequence of the gene coding for non-proteolytic
Clostridium botulinum type B neurotoxin: comparison with other
clostridial neurotoxins
Curr. Microbiol. 28 (2), 101-110 (1994)
94122659
REFERENCE
2 (bases 1 to 4051)
Hutson,R.A.
Direct Submission
Submitted (06-APR-1993) R.A. Hutson, AFRC Institute of Food
Research, Reading Laboratory, Microbiology Dept., Earley Gate,
Whiteknights Road, Reading, RG6 2EF, UK
Location/Qualifiers
1..4051
/organism="Clostridium botulinum"
/strain="Eklund 17b ATCC25765"
/isolate="type B"
/db_xref="taxon:1491"
/clone="PCBOTBI-8"
83..91
97..3972
/gene="botNT/B"
97..3972
/gene="botNT/B"
/codon_start=1
/transl_table=11
/product="BotNT/B"
/protein_id="CAA50482.1"
/db_xref="GI:296149"
/db_xref="SPTREMBL:Q08077"
/translation="MPVTINNFYNDPIDNDNIIMHEPPFARGTGTRYKAFKITDRIV
IIPERYTFGYKPEDKSGGIFNRDVCYIDPDYLTNDKKNIFPOTLIKLFNRKSK
PLGKLEMIINGIPYLGDRVPLEEFNTIASVTNKLISNPGVERKKGIFANLII
FGPGVLNENETIDIGIONHFASREGFGIMQMKFCPEVSVFVNVOENKKGASINRR
GYSPDALILMHELJHVLGLYGIKVDOLPIVPNEKKFFMOSTDITQAEELTFGQD
PSLISPTDKSIYDKVLQNFGRGIVRLKVLVCISDPNINIYKKNFKDKYKFVEDS
EGKYSIDVESFNKLYKSLMGLFTEINIAYENYKIKTRASTYSDSLPPVKIKLLDWEIY
TIEEFNISDKNMKGKQNAIKQAYEIEISKEHLAVYIQMCKSVKVPICIDVD
NENLFFIADKNSFSDLSKNERVEYNTQNNYIGNDFINELITDLTDLISKELPSENT
ESLTDNVDPVPEKOPAIKKVYTDENTIFQYLYSQTPLNTRDLSLTSSPDDALLVS
SKVYSFFSMDVYIKTANKVVEAGLWAGVKQIVDDPVEANKSSTMDKLTADISLIVPI
GLALNVGDETAKGNFSAFESLLEFIPELLIPVVGVELLESYIDNKNKIKRTI
DNALTRKVEKIDWYGLIVAQMLSTVTFYTIKEGMYKALNYQAALKEEIIKYKINI
YSEEKSNINIFNDINSKLANDINQAMNDINFTNECSVSLMKMPLAVKLLDF
DNTLKNNLTNDENKLYLIGSVEDEKSKVDKLTIIIPDLSTVNTSNIETIKFNKY
NSEILNNTILNLYRDNNLIDLSGYAKVEVDGVKLNQKOFKITSSADSKIRVTQN
QNIIFNSMFLDFSFWIRIPKYRNDIQNIHREYTIINCMKNNSCHKISIRGNRII
WTLDINGTKSVFYEINREDISEYINRFEYITINNDNAKITINGTLESNDMDIKD
IGEVTVNGEITFKLGDVDYDQFIMWKYSFNTQLNQSNIKETIKIOSYSEYLDKF
GNPLMYKYYFNAGKNKSYIKLVKDSVSGEILIRSKYNQNSNINIRNYLIGSKFI
IRRESNQSINDDIVRKEDYIHLDLVLHHEWRVYAYKYFKEDEKFLSLTSDSNEF
YKTIIEKEYDQPSYSCQLLPKKDESTDIGLIGIHRYFESGVLRKKYKDYFCISKW
```

YLKVRKKPKYKSLGCMNQFIPKDEGTE*

BASE COUNT 1679 a 384 c 645 g 1343 t
ORIGIN

alignment_scores:

Quality: 2099.00 Length: 438
Ratio: 4.962 Gaps: 0
Percent Similarity: 96.575 Percent Identity: 88.813

alignment_block:

US-09-910-186a-8 x CBBONTB ..

Align seg 1/1 to: CBBONTB from: 1 to: 4051

3 AsnLysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeuArgTy 19
|||||
2656 AATAAATAAATAGCGAAATTTTAAATAATATATTCTTAAATTTAAGATA 2705
19 rLysAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluV 36
|:::|
2706 TAGAGATAATAATTTAATAGATTTATCAGGATATGGAGCAAGGTAGAGG 2755
36 alTyrAspGlyValGluLeuAsnAspLysAsnGlnPheLysLeuThrSer 52
|||||
2756 TATATGATGGGTCAAGCTTAATGATGATAAAATCAATTTAAATTAAGT 2805
53 SerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePheAs 69
|||||
2806 TCAGCAGATAGTAAGATTTAGAGTCACTCAAAATCAGAATGATATATAT 2855
69 nSerValPheLeuAspPheSerValSerPheTrpIleArgIleProLysT 86
|:::|
2856 TAGTATGTTCCCTTGATTTTAGCGTTAGCTTTTGGATAAGGATACCTAA 2905
86 yLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIleIle 102
|:::|
2906 ATAGGAATGATGATATACAAAATTTATTTCAATAATGAATATACGATA 2955
103 AsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgGlyAsnAr 119
|||||
2956 AATTGATGAAAAATAATTCAGCGTGGAAAAATATCTATTAGGGGTAATAG 3005
119 gLileIleTrpThrLeuIleAspIleAsnGlyLysThrLysSerValPheP 136
|:::|
3006 GATAATATGGACCTTAATTGATATAAATGGAAAAACCAATCAGTATTT 3055
136 heGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTrpPhe 152
|||||
3056 TTCAATATAACATAAGAGAAGATATATCAGAGTATATAAATAGATGGTT 3105
153 PheValThrIleThrAsnAsnLeuAsnAsnAlaLysIleTyrIleAsnGl 169
|||||
3106 TTGTGAACATTACTAATAATTTGGATAATGCTAAATTTATATTAATGG 3155
169 yLysLeuGluSerAsnThrAspIleLysAspIleArgGluValIleAlaA 186
|:::|
3156 CAGGTAGAAATCAAAATATGATGATTAATAAGATATAGGAGAAGTTATT 3205
186 snGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThrGlnPhe 202
|||||
3206 ATGGTGAATAACATTTAAATTAGATGGTGGATAGATAGAACACAAATTT 3255
203 IleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGlnSerAs 219
|||||
3256 ATTGGATGAATATTTTATGATTTTAAATACCAATTAACCAATCAATCAA 3305
219 nIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeuLysAspP 236
|:::|
3306 TATTAAGAGATATATAAAATTCATCATATACCGAATACCTTAAAGATT 3355
236 heTrpGlyAsnProLeuMetTyrAsnLysLysGluTyrTyrMetPheAsnAla 252
|:::|

3356 TTTGGGAAATCCTTTAATGTATATAATAAAGAATATATATGTTTAAATCGG 3405
253 GlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProValGlyGl 269
|||||
3406 GGAATAAAAAATTCATATATTAACACTAGTGAAGAGATTCATCTGTAGGTGA 3455
269 uIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleAsnTyrA 286
|:::|
3456 AATATTAATACGTAGCAAAATAATCAGAATTCCAATTAATAATAAATATA 3505
286 rGAspLeuTyrIleGlyGluLysPheIleIleArgArgLysSerAsnSer 302
|:::|
3506 GAAATTTATATATGGAGAAAAATTTATTAAGAAGAGAGCAAAATTC 3555
303 GlnSerIleAsnAspIleValArgLysGluAspTyrIleTyrIleLysAs 319
|||||
3556 CAATCTAATAATGATGATATAGTTAGAAAAGAGATTATATACATCTAGA 3605
319 pPhePheAsnLeuAsnGlnGluTrpArgValTyrThrTyrLysTyrPheL 336
|:::|
3606 TTTGGTACTTCACCATGAAGAGCTGGAGAGTATATGCCTATAAATATTTTA 3655
336 ysLysGluGluLysLeuPheLeuAlaProIleSerAspSerAspGlu 352
|:::|
3656 AGGAACAGGAAGAAAAATTTGTTTTATCTATTATTAAGTGATCTTAATGAA 3705
353 LeuTyrAsnThrIleGlnIleLysGluTyrAspGluGlnProThrTyrSe 369
|:::|
3706 TTTTATAGACTATAGAAATAAAGAATATGATGAACACCATCATATAG 3755
369 rCysGlnLeuLeuPheLysLysAspGluSerThrAspGluIleGlyL 386
|:::|
3756 TTGTCAGTTGCTTTTAAAAAAGATGAAGAAAGTACTGATGATATAGGAT 3805
386 euIleGlyIleHisArgPheTyrGluSerGlyIleValPheGluGluTyr 402
|:::|
3806 TGATTGGTATTTCGTTTCTACGAATCTGGAGTTTACGTAAAAAGATAT 3855
403 LysAspTyrPheCysIleSerLysTrpTyrLeuLysGluValLysArgLy 419
|:::|
3856 AAAGATTATTTTGTATATAAGTAAATGGTACTTAAAGAGGTAAGAAAGAA 3905
419 sProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheIleProLysAspG 436
|:::|
3906 ACCATATAAGTCAAAATTTGGGATGTAATTTGGCAGTTTATTCTCTAAAGATG 3955
436 luGlyTrpThrGlu 440
|:::|
3956 AAGGGTGGAGCTGAA 3969

seq_name: gb_ba:CB0242628

seq_documentation_block:

LOCUS CB0242628 1326 bp DNA linear BCT 03-JUN-2000

DEFINITION Clostridium botulinum boNT/B gene, partial.

ACCESSION AJ242628

VERSION AJ242628.1 GI:4914467

KEYWORDS boNT/B gene; botulinum neurotoxin type B.

SOURCE Clostridium botulinum.

ORGANISM

Clostridium botulinum; Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae; Clostridium.

REFERENCE 1 (bases 1 to 1326)

AUTHORS Lalli,G., Herreros,J., Osborne,S.L., Montecucco,C., Rossetto,O. and Schiavo,G.

TITLE Functional characterisation of tetanus and botulinum neurotoxins binding domains

JOURNAL J. Cell. Sci. 112 (Pt 16), 2715-2724 (1999)

MEDLINE 99343691

REFERENCE 2 (bases 1 to 1326)

AUTHORS Schiavo,G.

TITLE Direct Submission

JOURNAL Submitted (26-MAY-1999) Schiavo G., Molecular Neuropathobiology,

Imperial Cancer Research Fund, 44 Lincoln s Inn Fields, WC2A 3PX,
UNITED KINGDOM

FEATURES

source

Location/Qualifiers

1..1326
/organism="Clostridium botulinum"
/db_xref="taxon:1491"
/note="natural mutant"

gene

1..1326
/gene="boNT/B"

CDS

<1..>1326
/gene="boNT/B"
/codon_start=3
/transl_table=11
/evidence=experimental
/product="botulinum neurotoxin type B"
/protein_id="CAB43706.1"
/db_xref="GI:4914468"
/db_xref="SPTREMBL:Q9X708"

translation="VTNNTILTEIFNKSKILNIIILNRYRDNNDLDSGYGANVE
VYDVELNDKNQFKLTSSNSTSEIRVTQNONIIFNSMFLDFSVEFWRIPKYKNDGION
YIHNEYTIINCINKNSGWKISIRGNRIIWLTDINGKTSVFEYSIREDDISDYINRW
FVTIYNSDNAKIYNGKLESNIDIKDGEVIANGELIFKLDGIDIDRTQFLWMKYFS
IFNTELSQNAKEIKYQISYELKDFWGNPLMYKEYFMNAGNKNKYIKLKDSSV
GELIRSKYNQNSNTINYRNLYIGKEFIIIRKNSQSINDDIVRKEDYIYLDFFNPNR
ENRVAYKDFKEEKLVLANIYDSNEFYKTIQIREYDEQPTYSQOLLPKKDEESTDE
IGLIGHREYESSIVLKDYNKVCISKWYKEVKRKPYNPNLGNWQFI"

BASE COUNT 573 a 105 c 201 g 447 t

ORIGIN

alignment_scores:

Quality: 2097.00 Length: 430
Ratio: 5.017 Gaps: 0
Percent Similarity: 97.209 Percent Identity: 91.395

alignment_block:

US-09-910-186a-8 x CBO242628 ..

Align seg 1/1 to: CBO242628 from: 1 to: 1326

3 AsnLysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeuArgTy 19
|||||
36 AACAAATATATACAAAATTTAAATAATATATATCTTAAATTTAAGATA 85
19 rLysAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluV 36
|||||
86 TAGGGATAAATAATTAATAGATTTATCAGGATATGGGCAATGTAGAGG 135
36 alTyrAspGlyValIleLeuAsnAspLysAsnGlnPheLysLysThrSer 52
|||||
136 TATATGATGGGTCGAGCTTAATGATATAAATCAATTTAAATTAACCTAGT 185
53 SerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePheAs 69
|||||
186 TCAACAACACAGCTAGATTAGAGTGACTGACACAAATCAGAAATATCATATTAA 235
69 nSerValPheLeuAspPheSerValSerPheTrpIleArgIleProLysT 86
|||||
236 TAGTATGTTCTTGATTTTAGTGTTAGCTTTTGGATAGAAATACCTAAAT 285
86 yrlYsAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIleIle 102
|||||
286 ATAAGATGATGGGTATACAAATTTATATTCATATGAATATACAAATTAAT 335
103 AsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgGlyAsnAr 119
|||||
336 AATTCGATTAATAAATAATCTCGCTGGGAAATATCTATTAGGGGTAATAG 385
119 gileIleTrpThrIleAspIleAsnGlyLysThrLysSerValPheP 136
|||||
386 GATAATATGGACTTAACTGATATAAATGGAAACCAATACGATGATTTT 435
136 heGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTrpPhe 152

|||||
436 TTGAATATAGCATAGAGACATATATCAGACTATATAAATAGATGGTTT 485
153 PheValThrIleThrAsnAsnLeuAsnAsnAlaLysIleTyrIleAsnGl 169
|||||
486 TTTGTAACATATACTATAATAATCGGATAACGCTAAAAATTTATTAATGG 535
169 yLysLeuGluSerAsnThrAspIleLysAspIleArgGluValIleAla 186
|||||
536 TAAGCTAGAAATCAATATAGATATATAAAGATATAGGAGAAGTTATGCTA 585
186 snGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThrGlnPhe 202
|||||
586 ATGCTGAAATTAATTTAAATTAGATGCTGATATAGATAGAACAAATTT 635
203 IleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGlnSerAs 219
|||||
636 ATTTGGATGAAATATTTTCACTATTTTAAACAGAAATTAAGTCAATCAAA 685
219 nIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLysLysAspP 236
|||||
686 TATTAAAGAAATATATAAATTCATCATATAGCGAATATTTAAAGATTT 735
236 heTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPheAsnAla 252
|||||
736 TTTGGGGAATCCCTTAAATGTACAAATAAAGAAATATATATCTTAAATCG 785
253 GlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProValGlyGl 269
|||||
786 GGAATAAAAAATTCATATATATAAATAAAGAAAGATTCATCTGTAGGTGA 835
269 uIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleAsnTyrA 286
|||||
836 AATTTTAACGCTAGCAAAATATAATCAAAATTCCAATTTATAAATATA 885
286 rGAspLeuTyrIleGlyGluLysPheIleIleArgArgLysSerAsnSer 302
|||||
886 GAAATTTATATATCGGAGAAAAATTTATATAGAAGAAAGTCAAAATCT 935
303 GlnSerIleAsnAspAspIleValArgLysGluAspTyrIleTyrLeuAs 319
|||||
936 CAATCTATAATGATGATATAGTTAGAAAAGAAGATATATATATATCTAGA 985
319 pPhePheAsnLeuAsnGlnGluTrpArgValTyrThrTyrLysTyrPheL 336
|||||
986 TTTTTCATTAATCAATCGAGAGTGGAGATATATGCCCTATAAAGATTTTA 1035
336 yLysGluGluGluLysLeuPheLeuAlaProIleSerAspSerAspGlu 352
|||||
1036 AGGAAGAGGAAAAAATTTGCTTTTAGCTAATATATATGATTCCTAATGAA 1085
353 LeuTyrAsnThrIleGlnIleLysGluTyrAspGluGlnProThrTyrSe 369
|||||
1086 TTTTACAAAACATATACAAATAAAGAAATATATGATGAACCAACATATAG 1135
369 rCysGlnLeuLeuPheLysLysAspGluGluSerThrAspGluIleGlyL 386
|||||
1136 TTGCAATTAATCTTTTAAAAAAGATGAGAAGACTGATGATGAGATAGGAT 1185
386 euIleGlyIleHisArgPheTyrGluSerGlyIleValPheGluGluTyr 402
|||||
1186 TGATTTGTTATTCATCGTTTTCACGAATCTGCAATTTGATTAATAAGATTA 1235
403 LysAspTyrPheCysIleSerLysTrpTyrLysLysGluValLysArgLy 419
|||||
1236 AAAAATTTATTTGTATAAAGTAAATGGTACTTAAAGAGGTAAGAAAGGAA 1285
419 sProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheIle 432
|||||
1286 ACCATATAAACCCCAATTTAGGGTGTAAATGGCAGTTTCATT 1325
seq_name: gb_pat:A69701

```

seq_documentation_block:
LOCUS      A59701                      3509 bp      DNA
DEFINITION Sequence 19 from Patent WO9807864.
ACCESSION  A59701
VERSION    A59701.1  GI:4774314
KEYWORDS
SOURCE     unidentified.
           unclassified.
REFERENCE  1 (bases 1 to 3509)
AUTHORS   Foster, K.A., Quinn, C.P. and Shone, C.C.
TITLE     RECOMBINANT TOXIN FRAGMENTS
JOURNAL   Patent: WO 9807864-A 19 26-FEB-1998;
          FOSTER KEITH ALAN (GB)
FEATURES
   source
       1..3509
           /organism="unidentified"
           /db_xref="taxon:32644"
   mat_peptide
       1..>3509
           /product="unnamed"
       1..>3509
           /note="unnamed protein product"
           /codon_start=1
           /protein_id="CAB42524.1"
           /db_xref="GI:4774315"
           /translation="MPVTINFNYNPDIIDNNIMMEPPFARGTGRTGRYYKAFKTIIDRIW
IIPERTFYGPEDFKNSGFIYRDCVEYDDPLNTNDKNKFLMTKMLKFRKRSK
PLGEKLEIMETIDIGYLDGRRVLEFNTINIASVTNKLISNPGEVERKKGGFANLII
FGPGVLNENETIDIGIQNHFSRGEFGGIMQMGPCPEYVSFNVQENKGASIFNRR
GYSPDALIMHLEIHLVHLCLYGIKVDDLPIDVNEKKFFMOSDLOAEELYTFEGOD
PSLIITPSTKSIYDKVLQNPGRGVDRNLKVLGCIISDPNININIIYKKFKDKYKEVEDS
EGYSIDVSEFSLKLYLMLPFGFTINIAENYKIKTRFASDLSPLPPKKNLDNIEY
TIEEGNIDSKOMEKEYRQONKAINQOAVEEISKEHLAVYKIQMCKSVKAPGICIDV
NEDLFTIADKNSFSDLSKNERIYNTQSNYIENDPFINELIIDLTLISKIELPSE
ESTUDFNVDVPVEKOPAIKKITFDGWTIFQYLSOTPLDIRDLSLTSFSDALLFS
NKVYSFSDMDYIKTANKVVEAGLFGWVKQIVNDFVEANKSNMTDKIADISLIVYI
GLALNVGNETAKGNFENAEIACASILLFEIPELLIPVVGAFLESYIDNKNKIILTI
DNALTRKNSQSDMYGLIYVAQWLSTVNTQYTIKEGMYKALNTQOALBEIILKYRNI
YSEKSRNSINDFNDNSKLEGINAQIDNINNFINGCSVSYLMKKMLPIALVEKLDNF
DITLKNLNIIDENKLYLIGSAEYKRYKVIKRTIMPFDLSITYNDTILIEFMFKY
NSPTILANIILNRYKKNLIDLSGYGAKVEYDGVELNDKNQFKLSSANSKTRVTRNI
NITLINSFVLDFSQFWIRPIKYKNDGIONYIHNEYTILNCKNNSGWKISIRGNRII
WTLIDNGKTSVFFVEYREDISEYINRWFFYITNLNNAKIYNGKLESFTDIDK
IRVIVANGELIFKLDGDIDQITFMWKYFISFNTFLESQSDIEERYKIQSYSEYKLOFV
GNPLMNKEYFMPNAGNKSNIYIKLKKDSPVGEILTRSKYNQNSKINYIRDLYIGEREKI
IRKNSQSDINDDIVRKEDYIYL"
1468 a      340 c      548 g      1153 t
BASE COUNT
ORIGIN

```

```

alignment_scores:
  Quality: 1667.00
  Ratio: 5.259
  Percent Similarity: 99.686
  Percent Identity: 99.686

alignment_block:
  US-09-910-186A-8 x A69701 ..
  Align seg 1/1 to: A69701 from: 1 to: 3509

1 MetAlaAsnLysTyrAsnSerGluIleLeuAsnAsnIleLeuAsnLe 17
  |||
  2554 ATGTTTAAATAAATAATAGCGAAATTTTAAATAATATTATCTTAAATTT 2603

17 aTgTyrLysAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysV 34
  |||
  2604 AAGATATATAGGATAATAATTTAATAGATTTATCAGGATATGGGGCAAGG 2653

34 aIGluValTyrAspGlyValGluLeuAsnAspLysAsnGlnPheLysLeu 50
  |||
  2654 TAGAGGTATATGATGGATCGAGCTTAATGATAAAAATCAATTTAAATTA 2703

51 ThrSerSerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleI 67

```

2704 ACTAGTTCAGCAAAATAGTAAGATTAGAGTGACTCAAAATCAGAATATCAT 2753
67 ePheAsnSerValPheLeuAspPheSerValSerPheTrpIleArgIleP 84
2754 ATTTAAATAGTGTGTCTTGGATTTAGCGTTAGCTTTGGATAAGAATAC 2803
84 rOLysTrYLysAsnAspGlyIleGlnAsnTrYLleHisAsnGluTrYLThr 100
2804 CTAATAATATAAGAAATGATGGTATACAAAATATATATTCATATGAATATACA 2853
101 lleIleAsnCYsMetLYsAsnAsnSerGlyTrpLYsIleSerIleAArgI 117
2854 AFAATTAATGTATGAAAAAATTAATTCGGCCCTGGAANAATATCTATTAGGG 2903
117 yAsnArgIlelleTrpThrLeuIleAspIleAsnGlyLYsThrLYsSerV 134
2904 TAATAGGATAATATGCAGCTTTAATTTGATATAAATGGAATAACCAATCGG 2953
134 alPhePheGluTrYLysAsnIleArgGluAspIleSerGluTrYLleAsnArg 150
2954 TATTTTTTGAATATAACATAGAAGAAGATATATCAGAGTATATAAATAGA 3003
151 TrpPhePheValThrIleThrAsnAsnLeuAsnAlaLYsIleTrYL 167
3004 TGGTTTTTGTAACTATTACTAATAATTTGAATAACGCTAAAATTTATAT 3053
167 eaSnGlyLYsLeuGluSerAsnThrAspIleLYsAspIleArgGluValI 184
3054 TAATGGTAAAGCTAGAAATCAAAATACAGATATTAAAGATATAGAAGAAGTTA 3103
184 leAlaAsnGlyCyllellePheLYsLeuAspGlyLYsAspIleAspArgThr 200
3104 TTGCTTAATGGTGAATAATATATTAATTAGATGGTGATATAGATAGAACA 3153
201 GlnPheIlelrPmetLYsTrYPheSerIlePheAsnThrGluLeuSerGI 217
3154 CAATTTATTTGGATGAAATATTTTCAGTATTTTAAATACGGAAATTAAGTCA 3203
217 sERAsnIleGluGluArgTrYLysIleGlnSerTrSerGluTrYLleuL 234
3204 ATCAAAATATTGAAGAAGATATAAATTCATCATATAGCGAATATTTAA 3253
234 yAspPheTrpGlyAsnProLeuMetTrYLysAsnLYsGluTrYLTrYrMetPhe 250
3254 RAAGATTTTGGGAAATCCTTTAATGTACATAAAGAATATTATATGCTTT 3303
251 AsnAlaGlyAsnLYsAsnSerTrYLleLYsLeuLYsLYsAspSerProVa 267
3304 AATCGGGGAAATAAAATTCATATATTAAACTAAAGAAGATTCACCTGT 3353
267 lGlyGluIleLeuThrArgSerLYsTrYLysAsnGlnAsnSerLYsTrYLleA 284
3354 AGGTCAAAATTTAACCGTACCAATATATCAAAATCTAAATATATAA 3403
284 snTrYArgAspLeuTrYLleGlyGluLYsPheIlelleArgArgLYsSer 300
3404 ATTTATAGAGATTTATATTGCGAAAAATTTATTTATAAAGAAGAAAGTCA 3453
301 AsnSerGlnSerIleAsnAspAspIleValArgLYsGluAspTrYLleTy 317
3454 AATTCCTCAATCTTAATGATGATATAGTTAGAAAAGAGATTTATATA 3503
317 rIleu 318
3504 TCTA 3507

```
seq_name: gb_pat:BD009886
seq_documentation_block:
LOCUS      BD009886
DEFINITION Recombinant t
ACCESSION  BD009886
```

PAT 31-JAN-2002

203 iletrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGlnSerAs 219
 6774 ATTGGATGAATATTTTCAGTGTCTTTTAAATACGGAATTAATCAATCAA 6823
 219 nilegluGluArgTyrLysIleGlnSerTyrSerGlu 231
 6824 TATTGAAGAAAGATATAAAATTCAATCATATAGCGAA 6860
 seq_name: gb_pat:AR0000029

seq_documentation_block:

LOCUS AR0000029 1330 bp DNA linear PAT 04-DEC-1998
 DEFINITION Sequence 22 from patent US 5736139.

ACCESSION AR000029

VERSION AR000029.1 GI:3962560

KEYWORDS

SOURCE

ORGANISM

Unclassified.

REFERENCE 1 (bases 1 to 1330)

AUTHORS Kink,J.A., Thalley,B.S., Stafford,D.C., Firca,J.R. and Padhye,N.V.

TITLE Treatment of Clostridium difficile induced disease

JOURNAL Patent: US 5736139-A 22 07-APR-1998;

FEATURES

source

1. .1330

/organism="unknown"

BASE COUNT 400 a 339 c 246 g 345 t

ORIGIN

alignment_scores:

Quality: 831.50 Length: 450

Ratio: 2.607 Gaps: 10

Percent Similarity: 70.889 Percent Identity: 40.000

alignment_block:

US-09-910-186a-8 x AR000029 ..

Align seg 1/1 to: AR000029 from: 1 to: 1330

4 LysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeuArgTyrLy 20
 28 GAATACATCAAGAACATCAATCAATCCATCCATCGTGAACCTGCGGTACGA 77
 20 sAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluValT 37
 78 ATCCAACTCACTGATCGACCTGCTCTCGCTGCTTCCAAATCAACATCG 127
 37 yAspGlyValGluLeuAsn.....AspLysAsnGlnPheLysLeuThr 51
 128 GTTCTAAAGTTAACTTCGATCCGATCGACAGAAATCAGATCCAGCTGTC 177
 52 SerSerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePh 68
 178 AATCTGGAATCTTCCAAATCGAAGTTATCTCTGGAAGTATGTCGTATA 227
 68 eAsnSerValPheLeuAspPheSerValSerPheTrpIleArgIleProL 85
 228 CAACTTAATGACGAAACTTCTCCACCTCTCTGATCCGTATCCCGA 277
 85 ystYrLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIle 101
 278 AATACTTCAAC.....TCCATCTCTCTGAAACAATGAATACACCATC 318
 102 IleAsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgGlyAs 118
 319 ATCAACTCATGGAAGAAATCTGTTGGAAGTATCTCTGAACATACGG 368
 118 nArgIleIleTrpThrLeuIleAspIleAsnGlyLysThrLysSerValP 135
 369 TGAATCATCTGGACTTCGAGGACACTCAGGAATCAACACGCGTGTG 418

135 hePheGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTrp 151
 419 TATTCAATACTCTCAGATCATCAATCTCTGACTACATCAATCGCTGG 468
 152 PhePheValThrIleThrAsnAsn...LeuAsnAsnAlaLysIleTyrIl 167
 469 ATCTTCGTTACCATCACCACAATCGTCTGAATAACTCCAAATCTACAT 518
 167 eAsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGluValI 184
 519 CAACGGCGCTGTATCGACCAACACCGATCTCCAAATCTGGGTAAACATCC 568
 184 leAlaAsnGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThr 200
 569 ACGCTTCTTAATAACATCATGTTCAAACTGGACGGTGTCTGCACACTCAC 618
 201 GlnPheIleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGl 217
 619 CGTACATCTGGATCAATACTTCAATCTGTTCGACAAAGAACTGAACGA 668
 217 nSerAsnIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeuL 234
 669 AAAAGAAATCAAGACCTGTACGACACACCAAGTCCAAATCTGGTATCCG 718
 234 ysAspPheTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPhe 250
 719 AAGACTTCTGGGTGACTACCTGCAGTACGACAAACCGTACTACTACATG 768
 251 AsnAlaGlyAsnLysAsnSerTyrIleLysLeuLysLys..... 263
 769 AATCTGTACGATCCGACAAATACGTTGACGTCACAAATAGTAGGTATCCG 818
 264AspSerProValGlyGluIleLeuThrArgSerL 275
 819 CGTTACATGTACCTGAAGGTCCGCTGTTCTGTTATGACTACCAACA 868
 275 ystYrAsnGlnAsnSerLysTyrIleAsnTyrArgAspLeuTyrIleGly 291
 869 TCTACCTGAACCTCTCC.....CTGTACCGTGGT 897
 292 GluLysPheIleIleArgLysSerAsnSerGlnSerIleAsnAspAs 308
 898 ACCAAATTCATCATCAAGAAATACGCGTCTGGTAAC.....AAGGACAA 941
 308 pIleValArgLysGluAspTyrIleTyrLeuAspPhePheAsnLeuAsnG 325
 942 TATCGTTCGCAACAATGATCGTGATATACATCAATGTTGTAGTAGAACA 991
 325 InGluTrpArgValTyrThrTyrLysTyrPheLysLysGluGluLys 341
 992 AAGAATACCGTCTGGCTACCAATGCTTCTCAGCGTGGTGTAGAAAAGATC 1041
 342 LeuPheLeuAlaProIleSerAspSerAspGluLeuTyrAsnThrIleGl 358
 1042 TTGCTCTGCTCTGGAATCCCGGACGCTGGTAAATCTCTCAGGTAGTTGT 1091
 358 nIleLysGluTyrAspGluGlnPro...ThrTyrSerCysGlnLeuLeuP 374
 1092 AATGAATCCAAAGAACGACGAGGTATCACTAACAATGCCAAATG.... 1137
 374 heLysLysAspGluSerThrAspGluIleGlyLeuIleGlyIleHis 390
 1138 ..AATCTCGAGGACAAACAATGTAACGATATCGGTTTCATCGTTTCCAC 1185
 391 ArgPheTyrGluSerGlyIleValPheGluGluTyrLysAspTyrPheCy 407
 1186 CAGTTCACCAATATCGCT.....AAACTGGT 1211
 407 sIleSerLysTrpTyrLeuLysGluValLysArgLysProTyrAsnLeuL 424
 1212 TGCCTCCAACTGGTACATCGTCAGATCGAAGCTTCC.....TCTCGCA 1255
 424 ysLeuGlyCysAsnTrpGlnPheIleProLysAspGluGlyTrpThrGlu 440

|||||||:||||:||||||| ||||:||||| |||
1256 CTCGGTTGCTCTGGGAGTTTCATCCCGGTTGATGACGGTTGGGTGAA 1305

OM of: US-09-910-186A-8 to: N_Geneseq_032802:* out_format : pfs
Date: Sep 2, 2002 4:04 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=frame1.p2n.model -DEV=xih
-Q=Cgn2_1/USPTO_spool/US09910186/runat_29082002.134726.16035/app_query.fasta_1.505
-DB=N_Geneseq_032802 -QFMT=fastcap -SUFFIX=ring -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-GAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-FGAPOP=6.000 -FGAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blomsum62
-TRANS=human40.cdi -LISP=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM-ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US09910186.@CGL_1_178 -NCPU=6 -ICPU=3 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -NO_XLPXY -WAIT -THREADS=1

Search information block:

Query: US-09-910-186A-8
Query length: 440
Database: N_Geneseq_032802:*
Database sequences: 1736436
Database length: 85845721
Search time (sec): 195.180000

score_list:

Sequence	Strd Orig	zScore	eScore	Len	! Documentation
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ87213 +	2350.00	3639.26	2.3e-19		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ87213 +	2350.00	3639.26	2.3e-19		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV30581 +	2340.00	3632.40	2.0e-19		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV30579 +	2333.00	3603.09	2.4e-19		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ87212 +	2227.00	3448.06	1.0e-17		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV30580 +	2102.00	3252.64	8.0e-17		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV26289 +	1667.00	2568.94	9.6e-13		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ87217 +	1168.50	1802.99	4.4e-92		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ87217 +	1168.50	1802.99	4.4e-92		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV30596 +	1168.50	1801.79	5.2e-92		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ87220 +	831.50	1279.45	6.4e-63		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ87216 +	831.50	1279.40	6.5e-63		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ87216 +	831.50	1279.38	6.5e-63		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAZ87216 +	831.50	1279.36	6.5e-63		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAZ87216 +	831.50	1279.36	6.5e-63		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ87216 +	831.50	1279.34	6.5e-63		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ87216 +	831.50	1279.30	6.6e-63		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV30576 +	831.50	1279.21	6.6e-63		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAZ87216 +	831.50	1278.87	6.9e-63		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV30572 +	831.50	1278.87	6.9e-63		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV30575 +	831.50	1277.97	7.8e-63		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ87218 +	831.50	1273.44	1.4e-62		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1996.DAT:AAZ87216 +	831.50	1269.50	2.3e-62		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ87216 +	831.50	1267.50	3.0e-62		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ87216 +	762.50	1172.11	6.1e-54		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT:AAZ87216 +	732.00	1124.98	2.6e-54		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1997.DAT:AAZ87216 +	732.00	1124.84	2.6e-54		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ87216 +	732.00	1124.83	2.6e-54		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ87216 +	732.00	1124.81	2.6e-54		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ87216 +	728.00	1118.59	5.9e-54		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV30583 +	693.50	1033.95	6.5e-51		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ87216 +	692.00	1032.40	2.9e-51		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ87216 +	688.50	1057.48	1.5e-50		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ87216 +	688.50	1056.64	1.7e-50		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ87216 +	688.50	1056.64	1.7e-50		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV30584 +	676.00	1036.81	2.1e-49		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV30591 +	615.50	942.75	3.7e-44		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAZ87216 +	612.50	938.70	6.1e-44		
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1991.DAT:AAZ87216 +	594.00	910.05	2.4e-43		

/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1999.DAT:AAZ87234 + 594.00 908.55 2.9e
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1987.DAT:AAV70345 + 594.00 907.21 3.5e
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV32580 + 594.00 907.17 3.5e
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1995.DAT:AAQ97490 + 594.00 900.82 7.9e
/SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA1994.DAT:AAQ57880 + 594.00 900.72 8.0e
seq_name: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/NA2000.DAT:AAZ87213
seq_documentation_block:
ID AAZ87213 standard; DNA; 1341 BP.
XX
AC AAZ87213;
XX
DT 08-MAY-2000 (first entry)
XX
DE DNA encoding synthetic BoNT serotype B (BoNTB) Hc fragment.
XX
KW Botulinum neurotoxin; heavy chain; BoNT; serotype B;
KW C-terminal fragment; Venezuelan equine encephalitis virus replicon;
KW VEE; botulism; vaccine; diagnosis; drug screening; ds.
XX
OS Clostridium botulinum.
OS Synthetic.
FH Key Location/Qualifiers
FT CDS 10..1333
FT /tag= a
FT /product= "Synthetic botulinum neurotoxin serotype B
PN (BoNTB) heavy chain C-terminal fragment (Hc)"
XX
XX WO200002524-A2.
XX
PD 20-JAN-2000.
XX
PF 09-JUL-1999; 99WO-US15570.
XX
PR 10-JUL-1998; 98US-0092416.
PR 12-MAY-1999; 99US-0133870.
XX
XX (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
XX
XX Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
WPI: 2000-160827/14.
XX P-PSDB; AAY77135.
XX
XX Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
toxin serotypes A-G, is used for inducing an immune response against
botulinum -
XX
XX Disclosure; Page 39-40; 54pp; English.
XX
XX The invention relates to novel vaccines that induce a protective immune
response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
and G (BoNTA-BoNTG). The vaccine of the invention is novel recombinant
DNA construct comprising a vector, and at least one nucleic acid
fragment comprising a C-terminal heavy chain fragment (Hc) from BoNT
serotypes A-G. In preferred embodiments of the invention, the vector is
a Venezuelan equine encephalitis virus (VEE) replicon vector. Use of
this vector results in the production of large amounts of a protein
encoded by a sequence cloned into the replicon. The constructs are used
to produce vaccines against botulism. The proteins can also be used as
diagnostic tools for the diagnosis of botulism. The transformed host
cells can be used to analyse the effectiveness of drugs and agents which
inhibit toxin effects. The vaccine currently used against botulism is
dangerous and expensive to produce, and contains formalin, which is very
painful for the recipient. Also, the vaccine is incomplete, in that only
5 of the 7 serotypes are represented in the formulation. The novel
vaccine overcomes these problems, as it is easily purified, and
available in large quantities. It is also expressed in the lymph nodes
for a better immune response. Sequences AAZ87212-287217 represent
synthetic DNA sequences encoding BoNT Hc fragments used in the present
invention. These were optimised for codon usage for expression in yeast.

xx

SQ Sequence 1341 BP; 442 A; 332 C; 235 G; 332 T; 0 other;

alignment_scores:

Quality: 2350.00 Length: 440
Ratio: 5.341 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:

US-09-910-186a-8 x AA287213

Align seg 1/1 to: AA287213 from: 1 to: 1341

1 MetAlaAsnLysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeu 17
10 ATGGCCACAATAACAATTCGAAATCTCGAACAATATCATCTCGAAGCT 59
17 uArgTyrLysAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysV 34
60 GGGTTACAAAGACAACAATCTGATCGATCTGTCTGGTTACGGTGCTAAAG 109
34 alGluValTyrAspGlyValGluLeuAsnAspLysAsnGlnPheLysLeu 50
110 TTGAAGTATACGACGGTGTGAAGTGAATGACAAAGACCAAGTTCAAACTG 159
51 ThrSerSerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleI 67
160 ACCTTCCGCTTAACCTTAAGATCGGTGTACTCAGAAATGAGAAATCAT 209
67 ePheAsnSerValPheLeuAspPheSerValSerPheTrpIleArgIleP 84
210 CTTCAACTCCGTATTCTCGGACTTCTCTGTTTCTCTTGGATTCGTATCC 259
84 roLysTyrLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThr 100
260 CGAAATACAAGAACACGGTATCCAGAAATTATACATCCACAATGAATACACC 309
101 IleIleAsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgG 117
310 ATCATCAACTGCATGGAAGAAATACCTCGTGGTGAAGATCTCCATCCGGG 359
117 yAsnArgIleIleTrpThrLeuIleAspIleAsnGlyLysThrLysSerV 134
360 TAACCGTATCATCTGGACTCTGATCGATATCAACGGTAAGACCAATCTG 409
134 alPhePheGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArg 150
410 TATTCTTCGAATACAAACATCCGTGAAGACATCTCTGAATACATCAATCGC 459
151 TrpPhePheValThrIleThrAsnAsnLeuAsnAlaLysIleTyrI 167
460 TGGTCTTCGTTACCATCACCATAACCTGAACATGCTAAATCTACAT 509
167 eAsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGluValI 184
510 CAACGGTAACTGGAATCTAATACCGACATCAAGACATCCGTGAAGTTA 559
184 leAlaAsnGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThr 200
560 TCGCTTAACGGTGAATATCATCTCAACCTGGACGGTGACATCGATCGTACC 609
201 GlnPheIleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerG 217
610 CAGTTCATCTGGATGAATFACTTCTCCATCTTCAACACCGAAGCTGTCA 659
217 nSerAsnIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeuL 234
660 GTCCAATATCGAAGAACGGTACAAGATCCAGTCTTACTCCGAATACCTGA 709
234 ysAspPheTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPhe 250
710 AAGATCTTCGGGGTAATCCGCTGTGTGTAACAACAAGAAATACTATATGTC 759

251 AsnAlaGlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProVa 267
760 AATGCTGTGTAAACAAGACTTTACATCAAACTGAAGAAAGACTCTCCGCT 809
267 lGlyGluIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleA 284
810 TGGTGAATCTCTGACTCGTTCCAATACAAACACAGAACTCTTAATACATCA 859
284 snTyrArgAspLeuTyrIleGlyGluLysPheIleIleArgArgLysSer 300
860 ACTACCGGACCTGTACATCGGTGAAAGATTTCATCCGTGCAATCT 909
301 AsnSerGlnSerIleAsnAspIleValArgLysGluAspTyrIleTyr 317
910 AACTCTCAGTCCATCAATGATGACATCGTACGTAAGAAGACTACATCTA 959
317 rLeuAspPhePheAsnLeuAsnGlnGluTrpArgValTyrThrTyrLysT 334
960 CTGGACTTCTTCAACCTGAATCAGGAATGGCGTGTATACACCTACAAGT 1009
334 yrPheLysLysGluGluLysLeuPheLeuAlaProIleSerAspSer 350
1010 ACTTCAAGAAAGAGAGAAAGCTTTTCTCGCTCCGATCTCTGATTCC 1059
351 AspGluLeuTyrAsnThrIleGlnIleLysGluTyrAspGluGlnProth 367
1060 GACGAACTTACACACCATCCAGATCAAGAAATACGACGACACGCCGAC 1109
367 rTyrSerCysGlnLeuLeuPheLysLysAspGluGluSerThrAspGluI 384
1110 CTACTCTTCCAGCTGCTGTTCAGAAAGATGAAGAAATCTACTACGAAA 1159
384 leGlyLeuIleGlyIleHisArgPheTyrGluSerGlyIleValPheGlu 400
1160 TCGGCTGATCGGTATCCACCGTTTCTACGAATCTGGTATCGTATTCGAA 1209
401 GluTyrLysAspTyrPheCysIleSerLysTrpTyrLeuLysGluVally 417
1210 GAATACAAGACTACTTCTGCATCTCCAAATGGTACTGAAGAAAGTTAA 1259
417 sArgLysProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheIleProL 434
1260 ACGCAAAACCGTACAACCTGAAACTGGTTGCAATGGCAGTTTCATCCGA 1309
434 ysAspGluGlyTrpThrGlu 440
1310 AAGCGAAGGTGGACCGAA 1329
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAA54485
seq_documentation_block:
ID AAA54485 standard; DNA; 1341 BP.
XX
AC AAA54485;
XX
DT 11-APR-2001 (first entry)
XX
DE Botulinism toxin heavy chain C-terminal coding sequence (serotype B).
XX Botulinism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection; ds.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
XX Key Location/Qualifiers
FH 10..1332
FT CDS /*tag= a
FT /product= H_C peptide fragment
XX
PN W0200067700-A2.

1110 CTACTCTGCCAGCTGCTGTTCAAGAAAGATGAAGAATCTACTGACGAAA 1159
384 leGlyLeuileGlyIleHisArgPheTyrGluSerGlyIleValPheGlu 400
1160 TCGGTCGTGATCGGATATCCACCGTTTCACGAATCTGGTATCGTATCGAA 1209
401 GluTyrLysAspTyrPheCysIleSerLysTyrTrpTyrLeuLysGluVally 417
1210 CAATACAAAGACTACTCTGCATCTCCAATGGTACCTGAAGGAAGTTAA 1259
417 sArgLysProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheIleProL 434
1260 ACCGAAACCGTACAACTGAACTGGGTTGCAATGGCAGTTCATCCGA 1309
434 ysAspGluGlyTrpThrGlu 440
1310 AAGACGAGGTGGACCGAA 1329

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV30581

seq_documentation_block:

ID AAV30581 standard; DNA; 1547 BP.

XX AAV30581;

DT 07-DEC-1998 (first entry)

DE Clostridium botulinum toxin B fragment C gene in pETHisB.

XX Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;
KW botulism; BotB; ds.

XX OS Clostridium botulinum serotype B Danish strain.

OS Synthetic.

XX Key Location/Qualifiers

FT CDS 108..1526

FT /*tag= a

XX WO9808540-A1.

XX 05-MAR-1998.

XX 28-AUG-1997; 97WO-US15394.

XX 28-AUG-1996; 96US-0704159.

XX (OPHI-) OPHIDIAN PHARM INC.

XX Thalley BS, Williams JA;

XX WPI; 1998-230234/20.

XX P-PSDB; AAW68394.

XX Host cell containing recombinant expression vector encoding
PT Clostridium botulinum type B or E toxin - useful to treat humans
PT and other animals at risk of intoxication with clostridial toxin

XX Example 35; Page 303-305; 428pp; English.

XX This is the DNA sequence of the Clostridium botulinum serotype B
CC (Danish strain) toxin fragment C gene contained in plasmid pETHisB.
CC The encoded fragment C polypeptide (see AAW68394) has a His-tagged
CC N-terminal extension. The vector was used to express native
CC (i.e. non-fusion) soluble C fragment in Escherichia coli host
CC cells. The invention relates to recombinant proteins derived from
CC C. botulinum toxins. Methods are provided which allow for the
CC isolation of soluble recombinant proteins free of significant
CC endotoxin contamination. Preferred hosts for production of
CC recombinant proteins are E. coli, insect cells and yeast cells.
CC The recombinant toxins are used as immunogens for the production
CC of vaccines and antitoxins that are useful in the treatment of
CC humans and animals at risk of intoxication with clostridial toxin.

XX
SQ Sequence 1547 BP; 634 A; 148 C; 263 G; 502 T; 0 other;
alignment_scores:
Quality: 2340.00 Length: 440
Ratio: 5.330 Gaps: 0
Percent Similarity: 99.773 Percent Identity: 99.545
alignment_block:
US-09-910-186a-8 x AAV30581 ..
Align seg 1/1 to: AAV30581 from: 1 to: 1547
1 MetAlaAsnLysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLe 17
204 ATGTTTAAATAATATATAGCGAAATTTTAAATAATATATATCTTAAATTT 253
17 uArgTyrLysAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysV 34
254 AAGATATAAGGATAATAATTTAATAGATTTATCAGGATATGGGCAAGG 303
34 alGluValTyrAspGlyValGluLeuAsnAspLysAsnGlnPheLysLeu 50
304 TAGAGGTATATGATGGAGTCGAGCTTAATGATAAAATCAATTTAAATTA 353
51 ThrSerSerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleI 67
354 ACTAGTTCAGCAATATAGTAAGATTTAGAGTGACTCAAAATCAGATATCAT 403
67 ePheAsnSerValPheLeuAspPheSerValSerPheTrpIleArgIleP 84
404 ATTTAATAGTGTCTCTTGTATTTAGCGTTAGCTTTTGGATAAGAATAC 453
84 roLysTyrLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThr 100
454 CTAATATAAGAATGATGGTATACAAAATTTATTCATTAATGAATATACA 503
101 IleIleAsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgG1 117
504 ATAATTAATGTTGTAATAATAATTCGGCTGGAAAATATCTATTAGGGG 553
117 yAsnArgIleIleTrpThrLeuIleAspIleAsnGlyLysThrLysServ 134
554 TAATAGGATAATATGGACTTTAATTTGATATAAATGGAACCAAAATCGG 603
134 alPhePheGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArg 150
604 TATTTTTTGAATATAACATAAGAAAGATATATCAGAGTATATAAATAGA 653
151 TrpPhePheValThrIleThrAsnAsnLeuAsnAlaLysIleTyrI1 167
654 TGGTTTTTGTAACTATTAATAATTTGAATTAACGCTAAAAATTTATAT 703
167 eAsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGluValI 184
704 TAATGGTAAGCTAGAAATCAAAATACAGATATTAAAGATATAAGAGAAGTTA 753
184 leAlaAsnGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThr 200
754 TTGCTAATGGTGAATAATAATTTAAATTTAGATGGTGATATAGATAGAAC 803
201 GlnPheIleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerG1 217
804 CAATTTATTTGGATGAATATTTTCAGTATTTTAAATACGGAATTAAGTCA 853
217 nSerAsnIleGluArgTyrLysIleGlnSerTyrSerGluTyrLeuL 234
854 ATCAAAATATTGAGAAAGATATAAAATTTCAATCATATACGCAATATTAA 903
234 ysAspPheTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPhe 250
904 AAGATTTTTGGGGAATCCCTTTAATGTACAAATAAGAAATATATATGTTT 953


```
2954 TATTTTTGAATATAACATAAGAGAAGATATATACAGAGTATATAAATAGA 3003
151 TrpPheValThrIleThrAsnAsnLeuAsnAlaLysIleTyrIle 167
|||||
3004 TGGTTTTTGTAACTATTAATAATTTGAATAACGCTAAATTTATAT 3053
167 eAsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGluValI 184
|||||
3054 TAATGGTAAGCTAGAAATCAATACAGATATTAAGATATAAGAGAGTTA 3103
184 leAlaAsnGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThr 200
|||||
3104 TTGCTAATGGTGAATAATATTTAAATAGATGTGATATAGATAGAACA 3153
201 GlnPheIleIlePheMetLysTyrPheSerIlePheAsnThrGluLeuSerGI 217
|||||
3154 CAATTTAATTTGGGATGAATATTTTCAGTATTTTAAATACGGAATTAAGTCA 3203
217 nSerAsnIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeuL 234
|||||
3204 ATCAAAATATTGAAGAAGATATAAAATTCATCATATAGCGAATATTAA 3253
234 ysAspPheTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPhe 250
|||||
3254 AAGATTTTGGGAAATCCTTTATGTACATAAAGAAATATTATATGTTT 3303
251 AsnAlaGlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProva 267
|||||
3304 AATGCGGGGAATAAAATTCATATATTAACTAAAGAAAGATTCACCTGT 3353
267 lGlyGluIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleA 284
|||||
3354 AGGTGAATTTTAAACACGTAGCAATATATCAAAATTCATAATATATAA 3403
284 snTyrArgAspLeuTyrIleGlyGluLysPheIleIleArgGlySer 300
|||||
3404 ATTATAGAGATTTTATATTGGGAAAAATTTATTATAAGAAAGAACTCA 3453
301 AsnSerGlnSerIleAsnAspAspIleValArgLysGluAspTyrIleTy 317
|||||
3454 AATTCCTCAATCTAATATGATATATAGTATAGTAAAGAAAGATATATATA 3503
317 rLeuAspPhePheAsnLeuAsnGlnGluTrpArgValTyrThrTyrLysT 334
|||||
3504 TCTAGATTTTTTTAAATTAATCAAGAGTGGAGAGTATATACCTATAAAT 3553
334 yrPheLysLysGluGluGluLysLeuPheLeuAlaProIleSerAspSer 350
|||||
3554 ATTTTAAAGAAAGGAGGAAAGAAATTTGTTTAGCTCTCTATAAGTGATCT 3603
351 AspGluLeuTyrAsnThrIleGlnIleLysGluTyrAspGluGlnProTh 367
|||||
3604 GATGAGTTTACATATACATAAATAAAGAAATATCATCAACAGCCAC 3653
367 rTyrSerCysGlnLeuLeuPheLysLysAspGluGluSerThrAspGluI 384
|||||
3654 ATATAGTTGTCACTGCTTTTAAAAAAGATGAAGAAAGTACTGATCAGA 3703
384 leGlyLeuIleGlyIleHisArgPheTyrGluSerGlyIleValPheGlu 400
|||||
3704 TAGGATGATGGGTATTCATCGITTCACGAATCTGGAATTTGATTTTGA 3753
401 GluTyrLysAspTyrPheCysIleSerLysTyrTyrLysGluVally 417
|||||
3754 GAGTATAAGATATTTTGTATAAAGTAATGGTACTTAAAGAGGTAA 3803
417 sArgLysProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheIleProL 434
|||||
3804 AAGGAAACCATATAAATTTAAATTTGGGATGATTAATGGCAGTTTATTCCTA 3853
434 ysAspGluGlyTyrThrGlu 440
|||||
3854 AAGATGAAGGGTGGACTGAA 3873
```

```
seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA2001A.DAT:AAA54589
seq_documentation_block:
ID AAA54589 standard; DNA; 1347 BP.
```

```
XX
AC AAA54589;
DT 11-APR-2001 (first entry)
XX
DE Sequence encoding botulinum toxin C fragment (serotype B).
XX
KW Botulinum; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection; ds.
XX
OS Synthetic.
OS Clostridium botulinum.
XX
PN WO200067700-A2.
XX
PD 16-NOV-2000.
XX
PF 12-MAY-2000; 2000WO-US12890.
XX
PR 12-MAY-1999; 99US-0133865.
PR 12-MAY-1999; 99US-0133866.
PR 12-MAY-1999; 99US-0133867.
PR 12-MAY-1999; 99US-0133868.
PR 12-MAY-1999; 99US-0133869.
PR 12-MAY-1999; 99US-0133873.
PR 29-JUL-1999; 99US-0146192.
XX
(PUSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
XX
PI Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
XX
WPI; 2001-016048/02.
XX
New nucleic acids encoding the carboxy- or amino-terminal portions of
the heavy chain of botulinum neurotoxin of serotype A-G, useful as
vaccine against botulinum
Example 8; Page 38-39; 73pp; English.
XX
Botulinum neurotoxins are translated as a single 150 kDa polypeptide
chain and then posttranslationally nicked, forming a dichain
consisting of a 100 kDa heavy chain and a 50 kDa light chain which
remain linked by a disulfide bond. Nucleic acids encoding the
carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
chain of botulinum neurotoxin (BoNT) can be used in recombinant
expression vectors and expressed in transformed cells to produce
peptide antigens useful for eliciting an immune response to give
protective immunity against botulinum neurotoxin, which causes
botulism. The nucleic acids are expressible in a recombinant
organism such as Escherichia coli or Pichia pastoris. The use
of recombinant nucleic acids are advantageous since it eliminates
the need to culture large quantities of hazardous toxin-producing
bacterium. Production yield from the genetically engineered product
is also high and cost of production is lower. The nucleic acids can
be derived from Clostridium botulinum serotypes A-G.
XX
SQ Sequence 1347 BP; 439 A; 337 C; 238 G; 333 T; 0 other;
```

```
alignment_scores:
Quality: 2227.00      Length: 438
Ratio: 5.108          Gaps: 0
Percent Similarity: 99.543 Percent Identity: 99.315
alignment_block:
US-09-910-186A-8 x AAA54589 ..
```

Align seg 1/1 to: AAA54589 from: 1 to: 1347

```
3 AsnLysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeuArgTy 19
|||||
10 AACAAATACAAATCCGAAATCCGTAACAAATATCATCTCGAACCTGGTTA 59
|||||
19 rLysAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluV 36
|||||
60 CAAAGACAACAATCTGATCGATCTGCTGGTTACGGTGCTAAAGTTGAAG 109
|||||
36 alTyrAspGlyValGluLeuAsnAspLysAsnGlnPheLysLeuThrSer 52
|||||
110 TATACACCGGTGTGTGAACCTGAATGACAAGAACCAAGTTCAAACATGAC 159
|||||
53 SerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIlePheAs 69
|||||
160 TCCGCTTAACCTTAAGATCCGTTACTCAGATCAGAACATCATCTTCAA 209
|||||
69 nSerValPheLeuAspPheSerValSerPheTyrPheIleArgIleProLys 86
|||||
210 CTCGGTATCTCTGACTTCTCTGTTCTCTGATCCGATCCCGAAAT 259
|||||
86 yrlYasnaspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIleIle 102
|||||
260 ACAAGAACGACGGTATCCGAATATACATCCACAATGAATACACCATCATC 309
|||||
103 AsnCysMetLysAsnAsnSerGlyTyrLysIleSerIleArgGlyAsnAr 119
|||||
310 AACTGATGAAGAATAACCTCTGGTTGGAAGATCTCCATCCGCGGTAAACGG 359
|||||
119 gIleIleTyrThrLeuIleAspIleAsnGlyLysThrLysSerValPheP 136
|||||
360 TATCATCTGGACTCTGATCGATATCAACGGTAAGACCAATCTGTATTCT 409
|||||
136 heGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTyrPhe 152
|||||
410 TCGAATACAAACATCCGCTGAAGACATCTCGAATACATCAATCGTGGTTC 459
|||||
153 PheValThrIleThrAsnAsnLeuAsnAsnAlaLysIleTyrIleAsnG 169
|||||
460 TTCGTTACCATCAACATACCTTGAACATGCTTAAATCTTACATCAACGG 509
|||||
169 yLysLeuGluSerAsnThrAspIleLysAspIleArgGluValIleAla 186
|||||
510 TAAACTGGAATCTAATACCGACATCAAGACATCCGCGAAGTTATCGCTA 559
|||||
186 snGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThrGlnPhe 202
|||||
560 ACGGTGAAATCATCTTCAAACTGGACGGTGACATCGATCGATCCCAAGTTC 609
|||||
203 IleTyrMetLysTyrPheSerIlePheAsnThrGluLeuSerGlnSerAs 219
|||||
610 ATCTGGATGAATACCTTCCATCTTCAACACCGCACTGTCTCAGTCCAA 659
|||||
219 nIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeuLysAsp 236
|||||
660 TATCGAAGAACGGTACAGATCCAGTCTTACTCCGAATACCTCAAGACT 709
|||||
236 heTyrGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPheAsnAla 252
|||||
710 TCTGGGGTAATCCGCTGATGTACAACAAAGAATACTATATGTTCAATGCT 759
|||||
253 GlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProValGlyG 269
|||||
760 GGTAACAAGAACCTTATACATCAAACTGAAGAAAGACTCTCCGGTTGGTGA 809
|||||
269 uIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleAsnTyrA 286
|||||
810 ATC.CTGACTCGTTCCAAATACAAACAGAACTCTTAATATCATCACTACC 858
|||||
286 rgAspLeuTyrIleGlyGluLysPheIleIleArgLysSerAsnSer 302
|||||
859 GCGACCTGTACATCGGTGAAAGTTTTCATCATCGTCCGCAAAATCTAACTCT 908
|||||
```

```
303 GlnSerIleAsnAspIleValArgLysGluAspTyrIleTyrLeuAs 319
|||||
909 CAGTCCATCAATGA.GACATCGTACGTAAAGAAGACTACATCTACCTGGA 957
|||||
319 pPhePheAsnLeuAsnGlnGluTrpArgValTyrThrTyrLysTyrPheL 336
|||||
958 CTTCCTTCAACCTGAATCA.GAATGGCGGTATATACACCTACAAGTACTTCA 1006
|||||
336 yLysLysGluGluLysLeuPheLeuAlaProIleSerAspSerAspGlu 352
|||||
1007 AGAAAGAAGAAGAAAGCTTTCTCGCTCCGATCTGTATCCGACGAA 1056
|||||
353 LeuTyrAsnThrIleGlnIleLysGluTyrAspGluGlnProThrTyrSe 369
|||||
1057 CTCTACAACACCATCCAGATCAAGAATAGCAGACAGACCGGACTACTC 1106
|||||
369 rCysGlnLeuLeuPheLysLysAspGluGluSerThrAspGluIleGlyL 386
|||||
1107 TTGCCAGCTGCTGTTCAAGAAGATGAAGAATCTACTGACGAAATCGGTC 1156
|||||
386 euIleGlyIleHisArgPheTyrGluSerGlyIleValPheGluGluTyr 402
|||||
1157 TGATCGGTATCCACCGTTTCTACGAATCTGGTATCGTATTCGAAGAATAC 1206
|||||
403 LysAspTyrPheCysIleSerLysTyrTyrLeuLysGluValLysArgLy 419
|||||
1207 AAAGACTC.TTCTGCATCTCCAATGGTACTGAGGAAGTTAAACGCAA 1255
|||||
419 sProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheIleProLysAspG 436
|||||
1256 ACGGTACAACCTGAACCTGGTTGCAATTGGCAGTTTCATCCCGAAAGACG 1305
|||||
436 luGlyTyrThrGlu 440
|||||
1306 AAGGTTGGACCGAA 1319
|||||
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseqn-emb1/NA1998.DAT:AAV30580

seq_documentation_block:

ID AAV30580 standard; DNA; 1526 BP.

XX AAV30580;

XX AC

XX 07-DEC-1998 (first entry)

XX DE Clostridium botulinum toxin B fragment C gene in pHisBotb.

XX KW Antitoxin; vaccine; neurotoxin; toxin B; intoxication; immunogen;

XX KW botulism; BotB; ds.

XX OS Clostridium botulinum serotype B strain Eklund 17B.

XX OS Synthetic.

XX FH Key

XX CDS Location/Qualifiers

XX FT 108..1526

XX FT /*tag= a

XX PN WO9808540-Al.

XX PD 05-MAR-1998.

XX PF 28-AUG-1997; 97WO-US15394.

XX PR 28-AUG-1996; 96US-0704159.

XX XX (OPHI-) OPHIDIAN PHARM INC.

XX XX Thalley BS, Williams JA;

XX DR WPI; 1998-230234/20.

XX DR P-PSDB; AAW68393.

XX XX

PT Host cell containing recombinant expression vector encoding
PT Clostridium botulinum type B or E toxin - useful to treat humans
PT and other animals at risk of intoxication with clostridial toxin
XX
PS Example 35; Page 300-302; 428pp; English.
XX
CC This is the DNA sequence of the Clostridium botulinum serotype B
CC toxin fragment C gene contained in plasmid pHisBotB. The encoded
CC fragment C polypeptide (see AAW68393) has a histidine-tagged
CC N-terminal extension. The vector was used to express native
CC (i.e. non-fusion) soluble C fragment in Escherichia coli host
CC cells. The invention relates to recombinant proteins derived from
CC C. botulinum toxins. Methods are provided which allow for the
CC isolation of soluble recombinant proteins free of significant
CC endotoxin contamination. Preferred hosts for production of
CC recombinant proteins are E. coli, insect cells and yeast cells.
CC The recombinant toxins are used as immunogens for the production
CC of vaccines and antitoxins that are useful in the treatment of
CC humans and animals at risk of intoxication with clostridial toxin.
XX
SQ Sequence 1526 BP; 622 A; 143 C; 259 G; 502 T; 0 other;

ignment_scores:

Quality: 2102.00 Length: 440
Ratio: 4.958 Gaps: 0
Percent Similarity: 96.364 Percent Identity: 88.636

alignment_block:

US-09-910-186a-8 x AAV30580 ..

Align seg 1/1 to: AAV30580 from: 1 to: 1526

1 MetAlaAsnLysTyrAsnSerGluIleLeuAsnAsnIleLeuAsnIle 17
|||||
204 ATGTTTAAATAAATAGCGAAATTTTAAATAATATATCTTAAATTT 253
17 uArgTyrLysAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysV 34
|||||
254 AAGATATAGAGATAATAATTAATAGATTATCAGGATATGGAGCAAGG 303
34 aGluValTyrAspGlyValGluLeuAsnAspLysAsnGlnPheLysLeu 50
|||||
304 TAGAGGTATATGATGGGTCAAGCTTAATGATAAAATCAATTTAAATTA 353
51 ThrSerSerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleI 67
|||||
354 ACTAGTTGAGCAGATAGTAGATTAGAGTCACTCAAAATCAGAATATTAT 403
67 ePheAsnSerValPheLeuAspPheSerValSerPheTrpIleArgIleP 84
|||||
404 ATTTAATAGTATGTTCTTGTATTTTACGTTAGCTTTTGGATAAGGATAC 453
84 rLysTyrLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThr 100
|||||
454 CTAATATAGGAATGATGATATACAAAATTTATTCATAATGAATATACG 503
101 IleIleAsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgG 117
|||||
504 ATAATTAATGATGAATAAATTCAGCGTGAATAATCTATTAGGGG 553
117 yAsnArgIleIleTrpThrLeuIleAspIleAsnGlyLysThrLysSerV 134
|||||
554 TAATAGGATAATATGACCTTAATGATATAAATGGAATAACCAATACAG 603
134 aPhePheGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArg 150
|||||
604 TATTTTTCATATAACATAAGAGAAGATATATCAGAGTATATAAATAGA 653
151 TrpPhePheValThrIleThrAsnAsnLeuAsnAlaLysIleTyrI 167
|||||
654 TGGTTTTTGTAACTATTACTAATAAATTTGGATAATGCTAAATTTATAT 703

167 eAsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGluValI 184
|||||
704 TAATGGCAGCTTGAAGTCAAAATATGATATTAAGATATAGGAAGTTA 753
184 leAlaAsnGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThr 200
|||||
754 TTGTTAATGGTGAATAACATTTAAATTTAGATGGTGTAGATAGAACA 803
201 GlnPheIleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerG 217
|||||
804 CAATTTATTTGGATGAAATATTTTAGTATTTTAAATACCAATTAATCA 853
217 nSerAsnIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeu 234
|||||
854 ATCAAAATTTAAAGAGATATATAAAATTTCAATCATATAGCGAATCTTAA 903
234 ysAspPheTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPhe 250
|||||
904 AAGATTTTGGGGAATCCTTTAATGTATAATAAAGAATATATATATGTT 953
251 AsnAlaGlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProVa 267
|||||
954 AATCGGGGAATAAAATTCATATATTAACTAGTCAAGATTCATCTGT 1003
267 lGlyGluIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleA 284
|||||
1004 AGGTGAATATTAATACGTAGCAAAATATAATCAGAAATTTCCAATTAATA 1053
284 snTyrArgAspLeuTyrIleGlyGluLysPheIleIleArgArgLysSer 300
|||||
1054 ATTATAGAAATTTATATATTTGGAGAAAATTTATTTATAAGAGAGATCA 1103
301 AsnSerGlnSerIleAsnAspAspIleValArgLysGluAspTyrIleTy 317
|||||
1104 AATCTCAATCTATAAATGATGATATAGTTAGAAAGAGATATATATACA 1153
317 rLeuAspPhePheAsnLeuAsnGlnGluTrpArgValTyrThrTyrLys 334
|||||
1154 TCTAGATTTTGGTACTTCACCATGAAGAGTGGAGAGTATATGCTATAAT 1203
334 yrPheLysLysGluGluLysLeuPheLeuAlaProIleSerAspSer 350
|||||
1204 ATTTAAGGAACAGGAGAAAATTTGTTTATCTATTATTAAGTGATTCT 1253
351 AspGluLeuTyrAsnThrIleGlnIleLysGluTyrAspGluGlnProTh 367
|||||
1254 AATGNAATTTTATAGACTATAGAAATAAAGAAATATGATGACACCATC 1303
367 rTyrSerCysGlnLeuLeuPheLysLysAspGluSerThrAspGluI 384
|||||
1304 ATATAGTTTGCAGTTGCTTTTAAAAAAGATCAAGAAAGTACTGATGATA 1353
384 leGlyLeuIleGlyIleHisArgPheTyrGluSerGlyIleValPheGlu 400
|||||
1354 TAGGATGATTGGTATTCATCTCTTCTACGAATCTGGAGTTTACGTAAA 1403
401 GluTyrLysAspTyrPheCysIleSerLysTrpTyrLeuLysGluVal 417
|||||
1404 AAGTATAAAGATTTATTTGTTATAGTAATAGGTACTTTAAAGAGGTAAA 1453
417 sArgLysProTyrAsnLeuLysLeuGlyCysAsnTrpGlnPheIleProL 434
|||||
1454 AAGAAACCATATAAGTCAAAATTTGGGATGTAATTTGGCAGTTTATTCTTA 1503
434 ysAspGluGlyTrpThrGlu 440
|||||
1504 AAGATGAAGGTGGACTGAA 1523

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV36289

seq_documentation_block:

ID AAV26289 standard; DNA; 3509 BP.

XX

AAV26289;
27-JUL-1998 (first entry)
Recombinant botulinum neurotoxin type B LH728/B encoding DNA.
Botulinum; recombinant; Clostridium botulinum; neurotoxin;
Immunogen; detection; tetanus; non-toxic; toxin; ds.
Synthetic.
OS Clostridium botulinum.
XX Key Location/Qualifiers
FH 1..3509
FT CDS /tag= a
FT /product= "LH728/B"
FT /note= "no stop codon given"
XX WO9807864-A1.
XX 26-FEB-1998.
XX 22-AUG-1997; 97WO-GB02273.
XX 13-DEC-1996; 96GB-0025996.
XX 23-AUG-1996; 96GB-0017671.
XX (MICR-) MICROBIOLOGICAL RES AUTHORITY.
XX (SPEY-) SPEYWOOD LAB LTD.
XX Foster KA, Quinn CP, Shone CC;
XX WPI: 1998-169168/15.
XX P-PSDB; AAV56017.
XX
XX Recombinant neurotoxin polypeptides - used to develop therapeutic
XX agents, immunogens or as non-toxic standards for the detection of
XX neurotoxins
XX
XX Disclosure; Page 87-91; 137pp; English.
XX
XX The present sequence encodes a recombinant neurotoxin protein from
XX the present invention. The present invention describes recombinant
XX neurotoxin proteins which comprise a first and second domain, where
XX the first domain is adapted to cleave one or more vesicle or
XX plasma-membrane associated proteins essential to exocytosis, and where
XX the second domain is adapted: (a) to translocate the protein into a
XX cell; (b) to increase the solubility of the protein compared to the
XX solubility of the first domain on its own, or (c) both to translocate
XX the protein into a cell and to increase the solubility of the protein
XX compared to the solubility of the first domain on its own, the protein
XX being free of clostridial neurotoxin (CN) and free of CN precursor that
XX can be converted into toxin by proteolytic action. The recombinant
XX proteins can be used as therapeutic agents for targeting cells
XX expressing a relevant substrate. The products can also be used as
XX immunogens and as non-toxic standards for the assessment and development
XX of in vitro assays for the detection of functional botulinum or tetanus
XX neurotoxins either in foodstuffs or in environmental samples.
XX
XX Sequence 3509 BP; 1468 A; 340 C; 548 G; 1153 T; 0 other;

alignment_scores:

Quality: 1667.00 Length: 318
Ratio: 5.259 Gaps: 0
Percent Similarity: 99.686 Percent Identity: 99.686

alignment_block:

US-09-910-186A-8 x AAV26289 ..

Align seg 1/1 to: AAV26289 from: 1 to: 3509

1 MetAlaAsnLysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeu 17

|||||
2554 ATGTTTAAATAATAATAGCGAAATTTTAAATAATAATAATCTTAAATTT 2603
17 uArgTyrLysAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysV 34
|||||
2604 AAGATATAAGGATATAATATTTTAAATAGATTTATCAGGATATGGGCAAGG 2653
34 alGluValTyrAspGlyValGluLeuAsnAspLysAsnGlnPheLysLeu 50
|||||
2654 TAGAGGTATATGATGGAGTCGAGCTTAATGATAAAATCAATTTAAATTA 2703
51 ThrSerSerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleI 67
|||||
2704 ACTAGTTCAGCAATAGTAAGATTAGAGTGACTCAAAATCAGAATATCAT 2753
67 ePheAsnSerValPheLeuAspPheSerValSerPheTrpIleArgIleP 84
|||||
2754 ATTTAAATAGTGTCTTCCTTGTATTTAGCTTTAGCTTTTGGATAAGAATAC 2803
84 rOLysTyrLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThr 100
|||||
2804 CTAATATAAGAATGATGGTATACAAATATATTCATAATGAATATATACA 2853
101 IleIleAsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgG 117
|||||
2854 ATAATTAATTCGTATGAAAAATAATTCGGCTGGAAAAATATCTATTAGGGG 2903
117 yAsnArgIleIleTrpThrLeuIleAspIleAsnGlyLysThrLysSerV 134
|||||
2904 TAATAGGATAATATGGACTTTTAAATGATATAATGGAACCAAAATCGG 2953
134 alPhePheGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArg 150
|||||
2954 TATTTTGAATAATAACATAAGAGAAGATATATCAGAGTATATAAATAGA 3003
151 TrpPhePheValThrIleThrAsnAsnLeuAsnAsnAlaLysIleTyrI 167
|||||
3004 TGGTTTTTGTACTATTAATTTGAATTAACGCTAAATTTATAT 3053
167 eAsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGluValI 184
|||||
3054 TAATGGTAAGCTAGAAATCAAAATACAGATATTTAAAGATATAAGAGAAGTTA 3103
184 leAlaAsnGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThr 200
|||||
3104 TTGCTAATGGTGAATATAATTTAAATTAGATGGTGATATAGATAGAACA 3153
201 GlnPheIleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerG 217
|||||
3154 CAATTTATTTGGATGAAATATTTTCAATTTTAAACGGAATTAAGTCA 3203
217 nSerAsnIleGluArgTyrLysIleGlnSerTyrSerGluTyrLeuL 234
|||||
3204 ATCAAAATTTGAAGAAAGATATAAAATTTCAATCATATAGCGAATATTTAA 3253
234 ysAspPheTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPhe 250
|||||
3254 AAGATTTTGGGGAATACCTTTAATGTACAATAAAGAAATATTTATATGTTT 3303
251 AsnAlaGlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerProVa 267
|||||
3304 AATCGGGGAATAAAATTCATATATTTAAACTAAAGAAAGATTCACCTGT 3353
267 lGlyGluIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrIleA 284
|||||
3354 AGGTGAAATTTTAAACACGTAGCAAAATATAATCAAAATTTCTAAATATATAA 3403
284 snTyrArgAspLeuTyrIleGlyGluLysPheIleIleArgArgLysSer 300
|||||
3404 ATTATAGAGATTATATATTTGGAGAAAAATTTATTTAAGAGAAAGATCA 3453
301 AsnSerGlnSerIleAsnAspIleValArgLysGluAspTyrIleTyr 317
|||||


```
731 TGAAGGACTCTCTGGGAAACCCACTGCTTACGACACCCAGTACTACCTG 780
XX
XX
250 PheAsnAlaGlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerPr 266
||||| ||| :||| ||||| :||| :||| :||| :||| :||| :||| :|||
781 TTCACACGAGGTATGCAGAACATCTACATCAAGTACTTCCCAAGGCCTC 830
XX
XX
266 oValcGlyGluIleuThrArgSerLysTyrAsnGlnAsnSerLysTyrI 283
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
831 CATGGTGAGACCCCTCGTACCACACTTCAACAGCCGCC.....A 874
XX
XX
283 leAsnTyrArgAspLeuTyrIleGlyGluLysPheIleIleArgArgLys 299
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
875 TCAACTACCAGAACCTTACTCGGTTCGCTTCATCATCAAGAAGGCC 924
XX
XX
300 SerAsnSerGlnSerIle....AsnAspAspIleValArgLysGluAspTy 315
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
925 TCCAACTCCCGTACATCAACAGCAGCAACATCGTCCGTGAGGGTGACTA 974
XX
XX
315 rIleTyrLeuAspPhePheAsnLeuAsnGlnGlu...TyrArgValTyrT 331
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
975 CATCTACTGAACATCGACACATCTCCGACGAGTCTCTACGGTGTCTACG 1024
XX
XX
331 hrTyrLysTyrPheLysGluGluLysLeuPheLeuAlaProIle 347
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1025 TCCTGGTCAACTCCCAAGGAGATCCAGACCCAGCTGTCTCTGGCCCAATC 1074
XX
XX
348 SerAspSerAspGluLeuTyrAsnThrIleGlnIleLysGluTyrAspG 364
:||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1075 AACGACGACCCCTACTCTTACGAGCTCTCGAGATCAAGAAGTACTACGA 1124
XX
XX
364 uGlnProThrTyrSerCysGlnLeuPheLysLysAspGluSerT 381
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1125 GAAGACCACTACAACTGTCTGATCCTGTGCGAGAGGAC.....A 1165
XX
XX
381 hrAspGluIleGlyLeuIleGlyLeuHisArgPheTyrGluSer...Gly 396
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1166 CCAAGACCTTCGGACTCTCGGTATCGGTAAAGTTCGTCGAAGGACTACGGT 1215
XX
XX
397 IleValPheGluGluTyrLysAspTyrPheCysTyrSerLysTrpTyrLe 413
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1216 TAGCTCTGGACACCTACGACAACTATCTCTGATCTCCAGTGGTACCT 1265
XX
XX
413 uLysGluValLysArgLysProTyrAsnLeuLysLeuGlyCysAsnTrpG 430
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1266 GCGTCGTATCTCCGAGACATCAACAGCTCGCTCTGGATGTAACTGGC 1315
XX
XX
430 lnPheIleProLysAspGluGlyTyrThrGlu 440
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
1316 AGTTTCATCCCGAGTCGACGAGGGTTGGACCGAG 1347
XX
XX
```

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA2001A.DAT:AAA54491

seq_documentation_block:

ID AAA54491 standard; DNA; 1368 BP.

XX AC AAA54491;

XX DT 11-APR-2001 (first entry)

XX DE Botulinism toxin heavy chain C-terminal coding sequence (serotype G).

KW Botulinism; toxin; neurotoxin; heavy chain; recombinant expression;
KW recombinant vector; antigen; immune response; vaccine; bacterium;
KW infection; ds.

XX OS Synthetic.

OS Clostridium botulinum.

XX FH Key Location/Qualifiers

FT CDS 10..1359

FT /tag= a

FT /product= H₃C peptide fragment

XX

```
PN WO20067700-A2..
XX
XX PD 16-NOV-2000..
XX
XX PF 12-MAY-2000; 2000WO-US12890..
XX
XX PR 12-MAY-1999; 99US-0133865..
XX PR 12-MAY-1999; 99US-0133866..
XX PR 12-MAY-1999; 99US-0133867..
XX PR 12-MAY-1999; 99US-0133868..
XX PR 12-MAY-1999; 99US-0133869..
XX PR 12-MAY-1999; 99US-0133873..
XX PR 29-JUL-1999; 99US-0146192..
XX
XX (USSA ) US ARMY MEDICAL RES & MATERIAL COMMAND.
PA
XX Smith LA, Byrne MP, Middlebrook JL, Lapenotiere H;
PI
XX WPI: 2001-016048/02..
XX P-PSDB; AAB04167..
XX
XX New nucleic acids encoding the carboxy- or amino-terminal portions of
PT the heavy chain of botulinum neurotoxin of serotype A-G, useful as
PT vaccine against botulism
XX
XX Claim 2; Fig 10a; 73pp; English.
XX
XX Botulinism neurotoxins are translated as a single 150 kDa polypeptide
CC chain and then posttranslationally nicked, forming a dichain
CC consisting of a 100 kDa heavy chain and a 50 kDa light chain which
CC remain linked by a disulfide bond. Nucleic acids encoding the
CC carboxy-terminal (HC) or amino-terminal (HN) portion of the heavy
CC chain of botulinum neurotoxin (BoNT) can be used in recombinant
CC expression vectors and expressed in transformed cells to produce
CC peptide antigens useful for eliciting an immune response to give
CC protective immunity against botulinum neurotoxin, which causes
CC botulism. The nucleic acids are expressible in a recombinant
CC organisms such as Escherichia coli or Pichia pastoris. The use
CC of recombinant nucleic acids are advantageous since it eliminates
CC the need to culture large quantities of hazardous toxin-producing
CC bacterium. Production yield from the genetically engineered product
CC is also high and cost of production is lower. The nucleic acids can
CC be derived from Clostridium botulinum serotypes A-G.
XX
XX Sequence 1368 BP; 366 A; 445 C; 267 G; 290 T; 0 other;
```

alignment_scores:

Quality: 1168.50 Length: 444

Ratio: 3.246 Gaps: 7

Percent Similarity: 81.081 Percent Identity: 50.000

alignment_block:

US-09-910-186A-8 x AAA54491 ..

Align seg 1/1 to: AAA54491 from: 1 to: 1368

```
3 AsnLysTyrAsnSerGluIleLeuAsnAsnIleIleLeuAsnLeuArgTy 19
||||| ||| :||| ||||| :||| ||||| :||| ||||| :||| ||
40 AACAACTACATCTCCACATCTCTCTCCAGGCCATCTGCTCCCTGTCTCTA 89
XX
XX 19 rLysAspAsnAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGlu 36
||||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
90 CCGTGGTGGTGGTGTGATGACTCTCTCCGGTTACGGAGCCACCATGAACG 139
XX
XX 36 alTyrAspGlyValGluLeuAsnAsp.....LysAsnGlnPheLysLeu 50
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
140 TCGTTCGCGAGCTCATCTTCAACGACATCGGTACGTCAGTTCAGCTG 189
XX
XX 51 ThrSerSerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleI 67
:||||| ||||| :||| :||| :||| :||| :||| :||| :||| :|||
190 AACAACTCCGAGAACTCCAACTATCCCGCCACCACCAAGTCCCAAGTTCGTCGT 239
```

```

67 ePheAsnSerValPheLeuAspPheSerValSerPheThrPheIleArgIleP 84
   :::::::::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|
240 CTAGGACTCCATGTTTCGACAACTTCTCCATCAACTTCTGGTCCGTACCC 289
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
84 roLysTyrLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThr 100
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
290 CAAAGTACAAACAACACGACATCCAGACCTACCTCGACAGACGAGTACACC 339
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
101 IleIleAsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgG 117
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
340 ATCACTCTCTGTATCAAGAACACCTCCGGTTCGAGGTCTCCATCAAGG 389
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
117 yAsnArgIleIleThrLeuIleAspIleAsnGlyLysThrLysSerV 134
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
390 AAACCGTATCATCTGGACCTCATCGACGTCACGCAACGCCAAGTCCA 439
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
134 aPhePheGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArg 150
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
440 TCTTCTTCGAGTACTCCATCAAGGACAACTCTCGGACTACATCAACAAG 489
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
151 TrpPhePheValThrIleThrAsnAsn...LeuAsnAsnAlaLysIleTy 166
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
490 TGGTCTCTCATCATCATCAACACGACCTCTGGGTAAACGCAACATCTPA 539
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
166 rIleAsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGluV 183
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
540 CATCAACGGTTCCTTCAGAAAGTCCGAGAGATCCTGAACCTGGACCGTA 589
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
183 aIleAlaAsnGlyGluIleIlePheLysLeuAspGlyAspIleAspArg 199
   :::::::::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|
590 TCAACTCTCTCCAAACGACATCGACTTCAAGCTGATCAACTGTACCGACACC 639
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
200 ThrGlnPheIleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSe 216
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
640 ACAAGTTCGCTGGATCAAGCACTTCAACATCTTCGGTCTGGTGAAGTAA 689
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
216 rGlnSerAsnIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrL 233
   :::::::::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|
690 CGCCACGAGGTCCTCCCTCTGACTGGATCCAGTCTCCACCAACACACC 739
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
233 euLysAspPheTrpGlyAsnProLeuMetTyrAsnLysGluTyrTrpMet 249
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
740 TGAAGGACTTCTGGGGAACCCACTCGGTTACGACACCCAGTACTACCTG 789
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
250 PheAsnAlaGlyAsnLysAsnSerTyrIleLysLeuLysLysAspSerPr 266
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
790 TTCACACGAGGTATGCAGAACATCTACATCAAGTACTCTCCAAAGCCTC 839
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
266 oValGlyGluIleLeuThrArgSerLysTyrAsnGlnAsnSerLysTyrI 283
   :::::::::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|
840 CATGGGTGAGACCGCCCTCGTACCAACTTCAACAACGCGCC.....A 883
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
283 leAsnTyrArgAspLeuTyrIleGlyGluLysPheIleIleArgLys 299
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
884 TCAACTACCAAGAACCTGTACCTCGGTCGCTTCATCATCAAGAAGGCC 933
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
300 SerAsnSerGlnSerIle...AsnAspIleValArgLysGluAspTy 315
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
934 TCCAACCTCCGTAACATCAACAACGACCAACATCTCGCGTGGGGTGACTA 983
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
315 rIleTyrLeuAspPhePheAsnLeuAsnGlnGlu...TrpArgValTyrT 331
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
984 CATCTACCTGAACTCGACAAACATCTCCGACGAGTCTTACCGTGTCTACG 1033
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
331 hrTyrLysTyrPheLysLysGluGluLysLeuPheIleAlaProIle 347
   :::::::::::::::::::::|::|::|::|::|::|::|::|::|::|::|::|
1034 TCCCTGGTCAACTCCAAAGGAGATCCAGACCCAGCTGTCTTCTGGCCCCAATC 1083
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
348 SerAspSerAspLeuTyrAsnThrIleGlnIleLysGluTyrAspG 364
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1084 AACGACGACCTTACCTTCTACGACGCTCTGCAGATCAAGAAGTACTACGA 1133
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
364 uGlnProThrTyrSerCysGlnLeuLeuPheLysLysAspGluGluSerT 381

```

```

::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1134 GAAGACCACTACAACTGTCTAGATCTGTGCGAGAAGGAC.....A 1174
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
381 hrAspGluIleGlyLeuIleGlyIleHisArgPheTyrGluSer...Gly 396
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1175 CCAAGACCTTCGGACTGTCGGTATCGGTAACTTCGTCGAAGGACTACGGT 1224
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
397 IleValPheGluGluTyrLysAspTyrPheCysIleSerLysTrpTyrIle 413
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1225 TACGTCTGGGACACCTACGACAACTACTCTGTATCTCCAGTGGTACCT 1274
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
413 uLysGluValLysArgLysProTyrAsnLeuLysLeuGlyCysAsnTrpG 430
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1275 GGGTCTGTCTCCGAGAACATCAACAAGCTGGTCTGGATGTAACCTGCG 1324
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
430 InPheIleProLysAspGluGlyTrpThrGlu 440
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
1325 AGTTCATCCAGTCGACGAGGGTTGGACCGAG 1356
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

seq_name: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/NA1998.DAT:AAV30596
seq_documentation_block:
ID   AAV30596 standard; DNA; 1535 BP.
XX
AC   AAV30596;
XX
DT   07-DEC-1998 (first entry)
XX
DE   Clostridium botulinum type G toxin C fragment gene in pETHsb.
XX
KW   Antitoxin; vaccine; neurotoxin; toxin G; intoxication; immunogen;
   botulism; BotG; ds.
XX
OS   Clostridium botulinum serotype G 113/30 strain.
XX
FH   Key Location/Qualifiers
FT   CDS 108..1529
FT     /*tag= a
XX
PN   WO9808540-A1.
XX
PD   05-MAR-1998.
XX
PF   28-AUG-1997; 97WO-US15394.
XX
PR   28-AUG-1996; 96US-0704159.
XX
PA   (OPHI-) OPHIDIAN PHARM INC.
XX
PI   Thalley BS, Williams JA;
XX
DR   WPI; 1998-230234/20.
DR   P-PSDB; AAW68400.
XX
PT   Host cell containing recombinant expression vector encoding
PT   Clostridium botulinum type B or E toxin - useful to treat humans
PT   and other animals at risk of intoxication with clostridial toxin
XX
PS   Example 49; Page 376-378; 428pp; English.
XX
CC   This is the DNA sequence of the Clostridium botulinum serotype G
CC   (113/30 strain) neurotoxin fragment C gene contained in plasmid
CC   pETHsb. The encoded BotG fragment C polypeptide (see AAW68400) has a
CC   His-tagged N-terminal extension. The vector can be used to express
CC   native (i.e. non-fusion) soluble C fragment in Escherichia coli host
CC   cells. The invention relates to recombinant proteins derived from
CC   C. botulinum toxins, especially type B and type E toxins. Methods
CC   are provided which allow for the isolation of soluble recombinant
CC   proteins free of significant endotoxin contamination. Preferred
CC   hosts for production of recombinant proteins are E. coli, insect
CC   cells and yeast cells. The recombinant toxins are used as
CC   immunogens for the production of vaccines and antitoxins that are

```


CC useful in the treatment of humans and animals at risk of
CC intoxication with clostridial toxin.
XX
SQ Sequence 1535 BP; 596 A; 172 C; 246 G; 521 T; 0 other;

```
alignment_scores:
  Quality: 1168.50
  Ratio: 3.246
  Percent Similarity: 81.081
  Length: 444
  Gaps: 7
  Percent Identity: 50.000
```

```
alignment_block:
US-09-910-186A-8 x AAV30596
```

Align seq 1/1 to: AAV30596 from: 1 to: 1535

3 AsnLysTyrAsnSerSerGluIleLeuAsnAsnIleIleLeuAsnLeuArgTyr 19
 210 AATAAATATATATAGTAATAATAGTAGTAATGCTATTTTAAAGTTAAAGTTA 259
 19 rLysAspAsnAsnLeuIleAspLeuSerGlyTyrGlyValAlaLysValGlu 36
 260 TAGAGTGGCGGTTAATAGATTATCTGGATATGGTGCAACTATGAATG 309
 36 alTyrAspGlyValGluLeuAsnAsp.....LysAsnGlnPheLysLeu 50
 310 TAGGTTCCAGATCTTATCTTAAATGATATAGGAATGTCGAATTTAAATTA 359
 51 ThrSerSerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIle 67
 360 AATAATCTCGAAATAGTAATAATATACGCCAGATCAAAATGTAATTCGTGT 409
 67 ePheAsnSerValPheLeuAspPheSerValSerPheTrpIleArgIleP 84
 410 ATATGATAGTATGTTGATAATTTTACCATTAACCTTTTGGGTGAAGNCTC 459
 84 roLysTyrLysAsnAspGlyIleGlnAsnTyrIleHisGlnGluTyrThr 100
 460 CTAAATATATAATATATGATATACAACTTATCTTCAAAATGAGTATACA 509
 101 IleIleAsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgG 117
 510 ATAAATTAGTTGTATAAAANAATGCTCAGGATGGAAAGTATCTATTAAAGG 559
 117 yAsnArgIleIleIleTrpThrLeuIleAspIleAsnGlyLysThrLysSer 134
 560 AAATAGATAAATAGCATTAATAGATGTTAAATGCAAAATCTAAATCAA 609
 134 alPhePheGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArg 150
 610 TATTTTTCGAATATAGTATAAAGAATAATATATCAGATTAATATAAATAA 659
 151 TrpPhePheValThrIleThrAsnAsn...LeuAsnAsnAlaLysIleTy 166
 660 TGGTTTTCCATAACTATTACTAATGATAGATTAGTAGTACGCCAAATATTA 709
 166 rIleAsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGlu 183
 710 TATAAATGGAGCTTTGAAAAAAGGTGAAAAATTTTAAACTTAGATAGAA 759
 183 alIleAlaAsnGlyGluIleIlePheLysLeuAspGlyLysAspArg 199
 760 TTAAATCTAGTAAATGATATAGACTCAAAATTAATTAATTTGTACAGATACT 809
 200 ThrGlnPheIleTrpMetLysTyrPheSerIlePheAsnThrGluLeu 216
 810 ACTAAATTTGTTTGGATTAAAGGATTTTAAATTTTGGTAGAGAATATAA 859
 216 rGlnSerAsnIleGluGluArgTyrLysIleGlnSerTyrSerGluTyr 233
 860 TGCATCAAGATCTCTTCACTATATGGATTCAATCATCTACAAATACT 909
 233 euLysAspPheTrpGluVAsnProLeuMetTyrAsnLysGluTyrTyrMet 249

|||||
910 TAAAGATTTTGGGGAATCCTTTAAGATACGATACACAACTACTATCTG 959
250 PheAsnAlaGluAsnLysAsnSerTyrrileLysLeuLysLysAspSer 266
|||||
960 TTAATCAAGGTATGCCAAATATCTATATAAAGTATTTTAGTAAGCTTC 1009
266 ovalGluLeuLeuThrArgSerLysTyrrAsnGlnAsnSerLysTyrr 283
|||||
1010 TATGGGGAAACTGCACCACCGCTACAAACTTTAAATAATGCCAGCA. . . .A 1053
283 leAsnTyArgAspLeuTyrrileGluLysPheIlelleArgAtgLys 299
|||||
1054 TAAATATCAAAATTTATATCTTGTTTACAGTTATTTATAAAAAAGCA 1103
300 SerAsnSerGlnSerIle. . .AsnAspIleValArgLysGluAspTy 315
|||||
1104 TCAAAATCTCGAATATAAATAATGATAATAATAGTCAGAGAAGGAGATTA 1153
315 rIleTyrrLeuAspPhePheAsnLeuAsnGlnGlu. . .TrpArgValTyrr 331
|||||
1154 TATATATCTTAATATGATAATATCTTGATGAATCTTACAGAGTATATG 1203
331 hrTyrrLysTyrrPheLysLysGluGluGluLysLeuPheLeuAlaProIle 347
:: |||||
1204 TTTTGGTGAATCTAAAGAAATTCAAACTCAATATTTTAGCACCACATA 1253
348 SerAspSerAspGluLeuTyrrAsnThrIleGlnIleLysGluTyrrAspG 364
:: |||||
1254 AATGATGATCTCAGCTTCTATGATGTACTACAAATAAAAAAATATTATGA 1303
364 uGlnProThrTyrrSerCysGlnLeuLeuPheLysLysAspGluGluSer 381
1304 AAAACAACATATTAATGTGAGATACTTGGCAAAAAGAT.A 1344
381 hrAspGluIleGlyLeuIleGlyIleHisArgPheTyrrGluSer. . .Gly 396
|||
1345 CTAAACATTTTGGCTGTTGGAAATTTGGTAATTTGTAATAAGATTATGGA 1394
397 lleValPheGluGluTyrrLysAspTyrrPheCysIleSerLysTyrrLe 413
|||||
1395 TATGTTTGGGATACCTATGATAATATTTTGGCAATAAGTCAGTGGTATCT 1444
413 uLysGluValLysArgLysProTyrrAsnLeuGluLysCysAsnTrpG 430
|||
1445 CAGAAGATATCTGAAAATATATAATAATTAAGTTGGGATGTAATTGGC 1494
430 lnPheIleProLysAspGluGlyTrpThrGlu 440
1495 AATTCATTCCCGTGGATCGAAGGATGGACAGAA 1526

seq_name: /SIDS1/qcadata/hold-qeneseq/qeneseq-emb1/NA2000.DAT:AAZ87220

seq_documentation block:

seq_documentation_block.
ID AAZ87220 standard; DNA; 1317 BP.

XX
XX
077/0777

AAZ87220;

XX
XX

DT 08-MAY-2000 (first entry)

XX
XX

DE DNA encoding native BONT s

XX

KW Botulinum neurotoxin; heavy

KW C-terminal fragment; Hc; V

KW VEE; botulism; vaccine; di

XX

OS Clostridium botulinum.

XX

FH	Key	Location/Q
FM	anc	1 1317

FT	CDS	1..1317
FM		/4+28= 3

```

ET      /*tag= a
ET      /product=

```

$$\frac{E_T}{E_M} = \frac{\text{product}}{\text{reactant}}$$

THE

```

FT XX /transl_except= (pos:1294..1314, aa:Gly)
PN WO200002524-A2.
XX
XX
PD 20-JAN-2000.
XX
XX 09-JUL-1999; 99WO-US15570.
XX
XX 10-JUL-1998; 98US-0092416.
PR 12-MAY-1999; 99US-0133870.
XX
XX (USME-) US MEDICAL RES INST INFECTIOUS DISEASES.
PA
XX
XX Lee JS, Pushko P, Smith JF, Parker M, Dertzbaugh MT, Smith L;
PI
XX WPI; 2000-160827/14.
DR
DR P-PSDB; AAY77142.
XX
XX Novel Botulinum neurotoxin vaccine comprising a fragment from botulinum
PT toxin serotypes A-G, is used for inducing an immune response against
PT botulinum -
XX
XX Example 3; Page 52; 54pp; English.
XX
XX The invention relates to novel vaccines that induce a protective immune
XX response against botulinum neurotoxin (BoNT) serotypes A, B, C, D, E, F
XX and G (BONTA-BONTG). The vaccine of the invention is novel recombinant
XX DNA construct comprising a vector, and at least one nucleic acid
XX fragment comprising a C-terminal heavy chain fragment (Hc) from BoNT
XX serotypes A-G. In preferred embodiments of the invention, the vector is a
XX Venezuelan equine encephalitis virus (VEE) replicon vector. Use of this
XX vector results in the production of large amounts of a protein encoded by
XX a sequence cloned into the replicon. The constructs are used to produce
XX vaccines against botulinum. The proteins can also be used as diagnostic
XX tools for the diagnosis of botulinum. The transformed host cells can be
XX used to analyse the effectiveness of drugs and agents which inhibit toxin
XX effects. The vaccine currently used against botulinum is dangerous
XX and expensive to produce, and contains formalin, which is very painful
XX for the recipient. Also, the vaccine is incomplete, in that only 5 of
XX the 7 serotypes are represented in the formulation. The novel vaccine
XX of overcomes these problems, as it is easily purified, and available in
XX large quantities. It is also expressed in the lymph nodes for a better
XX immune response. The present sequence represents DNA encoding native
XX BONTA heavy chain C-terminal fragment (Hc) used in an exemplification of
XX the present invention.
XX
XX Sequence 1317 BP; 560 A; 108 C; 211 G; 438 T; 0 other;
SQ

```

```

alignment_scores:
  Quality: 831.50      Length: 450
  Ratio: 2.607         Gaps: 10
Percent Similarity: 70.889 Percent Identity: 40.000

```

```
alignment_block:

```

```
US-09-910-186A-8 x AA287220 ..
```

```
Align seg 1/1 to: AA287220 from: 1 to: 1317
```

```

4 LysTyrAsnSerGluLeuAsnAsnIleLeuLeuAsnLeuArgTyrIly 20
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
28 GAATATATTAAAGAAATATTATTAACTCTCTATATTGAATTTAAGATATGA 77
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
20 sAspAsnLeuIleAspLeuSerGlyTyrGlyAlaLysValGluValT 37
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
78 AGTAAATCAATTAAGACTATCTAGGTAGTCATCAAAATAAATATG 127
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
37 yrAspGlyValGluLeuAsn.....AspLysAsnGlnPheLysLeuThr 51
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
128 GTAGTAAAGAAATTTTGATCCAAATAGATAAAATCAAAATCAATTTT 177
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
52 SerSerAlaAsnSerLysIleArgValThrGlnAsnGlnAsnIleIleP 68
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```

```

178 AATTTAGAAAGTAGTAAATTCAGGTAATTTTAAAAAATGCTATTGTATA 227
68 eAsnSerValPheLeuAspPheSerValSerPheTrpIleArgIleProL 85
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
228 TAATAGTATGTAGAAATTTTAGTACTAGCTTTTGGGATAAGAAATTCCTA 277
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
85 ysTyrLysAsnAspGlyIleGlnAsnTyrIleHisAsnGluTyrThrIle 101
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
278 AGTATTTTAAC.....AGTATAAGCTTAATAATAATGAATATACATA 318
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
102 IleAsnCysMetLysAsnAsnSerGlyTrpLysIleSerIleArgGlyAs 118
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
319 ATAAATTCATGGAAATAAATTCAGGATGGAAGATATCACTTAATATGG 368
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
118 nArgIleIleTrpThrIleAspIleAsnGlyLysThrLysSerValP 135
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
369 TGAATAATCTGGACTTTACAGGATACCTCAGGAAATAAAACAAGAGTAG 418
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
135 hePheGluTyrAsnIleArgGluAspIleSerGluTyrIleAsnArgTrp 151
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
419 TTTTAAATACAGTCAAAATGATTATATATCAGATTATATAACAGATGG 468
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
152 PhePheValThrIleThrAsnAsn....LeuAsnAsnAlaLysIleTyrIl 167
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
469 ATTTTGTAACTATCACATAATAATAGATTAAATACTCTAAAAATTTATAT 518
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
167 eAsnGlyLysLeuGluSerAsnThrAspIleLysAspIleArgGluValI 184
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
519 AAATGGAAGATTATAGATCAAAACCAATTTCAAAATTTAGGTAAATATTC 568
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
184 leAlaAsnGlyGluIleIlePheLysLeuAspGlyAspIleAspArgThr 200
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
569 ATGCTAGTAATAATAATAATGTTTAAATTAGATGGTTGTAGAGATACACAT 618
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
201 GlnPheIleTrpMetLysTyrPheSerIlePheAsnThrGluLeuSerGl 217
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
619 AGATATATTTGGATAAAATATTTAAATCTTTTGAAGGAATTAATAATGA 668
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
217 nSerAsnIleGluGluArgTyrLysIleGlnSerTyrSerGluTyrLeuL 234
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
669 AAAGAAATCAAGATTTATATGATAATCAATCAAAATTCAGGTATTTTAA 718
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
234 ysAspPheTrpGlyAsnProLeuMetTyrAsnLysGluTyrTyrMetPhe 250
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
719 AAGACTTTTGGGGTGATTTATTACAATATGATAAACCACATACATATATG 768
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
251 AsnAlaGlyAsnLysAsnSerTyrIleLysLeuLysLys..... 263
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
769 AATTTATATGATCCAAATAATATGTCGATGTAATAATAGTAGGTATTAG 818
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
264 .....AspSerProValGlyGluIleLeuThrArgSerL 275
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
819 AGGTATATATGATCTTAAAGGCCCTAGAGGTAGCGTAATGACTACAAACA 868
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
275 ysTyrAsnGlnAsnSerLysTyrIleAsnTyrArgAspLeuTyrIleGly 291
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
869 TTTATTTAAATTCAGT.....TTGTATAGGGGG 897
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
292 GluLysPheIleIleArgArgLysSerAsnSerGlnSerIleAsnAspAs 308
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
898 ACAAAATTTATTATAAAAAAATATGCTTCTGGAAAT.....AAAGATAA 941
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
308 pIleValArgLysGluAspTyrIleTyrLeuAspPhePheAsnLeuAsnG 325
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
942 TATTCTTAGAAATAATGATCGTGTATATATTAATGCTAGTCTAAATAATA 991
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
325 lnGluTrpArgValTyrThrTyrLysTyrPheLysLysGluGluLys 341
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
992 AAGATATAGTTAGTACTACTAATGTCATCACAGCGAGCGGTAGAAAAATA 1041
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
342 LeuPheLeuAlaProIleSerAspSerAspGluLeuTyrAsnThrIleGl 358
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::
1042 CTAAGTCGATTAGAAATACCTGATGCTAGGAATCTAAGTCAAGTAGTAGT 1091
   :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::

```


1042 TTGTCGTCTCTGGAATCCCGACGTTGGTAATCTGTCTCAGGTAGTTGT 1091
358 nileLysGluTyrAspGluInPro...ThrTyrSerCysGlnLeuLeup 374
1092 AATGAATCCAGACGACCGAGGTATCTACTACAAATCCAAATG... 1137
374 helLysLysAspGluGluSerThrAspGluIleGlyLeuIleHis 390
1138 ...AATCTGCAGCACACAAATGTTACGATATCGGTTTCATCGGTTCCAC 1185
391 ArgPheTyrGluSerGlyIleValPheGluGluTyrLysAspTyrPheCy 407
1186 CAGTTCACAAATATCGCT.....AAACTGCT 1211
407 sileSerLysTyrTyrLeuLysGluValLysArgLysProTyrAsnLeuL 424
1212 TGCCTTCCAACTGGTACATCGTCAGATCGACGTTCC.....TCTCGCA 1255
424 ysLeuGlyCysAsnTrpGlnPheIleProLysAspGluGlyTrpThrGlu 440
1256 CTTCTGGGTTGCTCTGGGAGTTTCATCCCGGTTGATGACGGTTGGGGTGAA 1305

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: September 2, 2002, 16:00:43 ; Search time 1859.46 Seconds
(without alignments)
15091.739 Million cell updates/sec

Title: US-09-910-186A-7

Perfect score: 1341

Sequence: 1 gaattcacgatggccaacaa.....ggaccgaatagtaagaattc 1341

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues

Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl:*

1: gb_ba:*

2: gb_hgt:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vi:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vi:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htgo_inv:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
------------	-------	-------------	--------	-------	-------------

1	742.8	55.4	4041	1	CLOBOTB	M81186 Clostridium
2	706.4	52.7	3876	1	AF295926	AF295926 Clostridi
3	706.4	52.7	3876	1	AF300465	AF300465 Clostridi
4	706.4	52.7	3876	1	AF300466	AF300466 Clostridi
5	706.4	52.7	11170	1	CBY13630	Y13630 Clostridium
6	696	51.9	3869	1	AF300467	AF300467 Clostridi
7	696	51.9	3869	1	AF300468	AF300468 Clostridi
8	696	51.9	3869	1	AF300469	AF300469 Clostridi
9	681.4	50.8	1326	1	CB0242628	AJ242628 Clostridi
10	663.4	49.5	4051	1	CB0242628	AJ242628 Clostridi
11	522.4	39.0	3509	6	A69701	A69701 Sequence 19
12	522.4	39.0	3509	6	BD009886	BD009886 Recombina
13	341	25.4	6862	1	CBNTNHTB	X87849 C.botulinum
14	263	19.6	1330	6	AR000029	AR000029 Sequence
15	263	19.6	1330	6	AR169140	AR169140 Sequence
16	263	19.6	1330	6	AX036243	AX036243 Sequence
17	263	19.6	1338	12	XX022962	U22962 Synthetic b
18	263	19.6	1402	6	AR000030	AR000030 Sequence
19	263	19.6	1402	6	AR169141	AR169141 Sequence
20	263	19.6	1402	6	AX036246	AX036246 Sequence
21	242.6	18.1	1313	6	A58946	A58946 Sequence 6
22	241	18.0	1299	12	AF251281	AF251281 Synthetic
23	224.8	16.8	1084	1	CBPPTOT	X70817 C.botulinum
24	220.2	16.4	3937	1	CBPPTOT	X74162 C.botulinum
25	212.6	15.9	1084	1	CBPPTOT	X70814 C.botulinum
26	212.6	15.9	1084	1	CBPPTOT	X70819 C.botulinum
27	135.4	10.1	4067	1	CBNTQXA	X73423 C.botulinum
28	134.2	10.0	3712	6	A49987	A49987 Sequence 4
29	131.4	9.8	1359	6	I28431	I28431 Sequence 3
30	129.4	9.6	4366	6	A42484	A42484 Sequence 12
31	129.2	9.6	3754	6	A37074	A37074 Sequence 17
32	129.2	9.6	3754	6	A42478	A42478 Sequence 6
33	129.2	9.6	3769	6	A37075	A37075 Sequence 18
34	129.2	9.6	3769	6	A49988	A49988 Sequence 5
35	129.2	9.6	4378	6	A42481	A42481 Sequence 9
36	123	9.2	3891	6	AR000031	AR000031 Sequence
37	123	9.2	3891	6	AR169142	AR169142 Sequence
38	123	9.2	3891	6	AX036248	AX036248 Sequence
39	123	9.2	4292	1	CBOTAG	X52066 Clostridium
40	123	9.2	4835	1	CLONEUR	X30196 C.botulinum
41	103.2	7.7	1766	6	A49989	A49989 Sequence 6
42	96	7.2	1293	6	A58945	A58945 Sequence 5
43	96	7.2	4199	1	CLOBONT	I35496 Clostridium
44	96	7.2	4209	1	CBPPTOT	X81714 C.botulinum
45	92.6	6.9	3952	1	CBNTTD	X54254 Clostridium

ALIGNMENTS

RESULT	1	CLOBOTB	4041 bp	DNA	linear	BCT 26-APR-1993
LOCUS		Clostridium botulinum neurotoxin type B (botB) gene, complete cds.				
DEFINITION		Clostridium botulinum neurotoxin type B (botB) gene, complete cds.				
ACCESSION		M81186				
VERSION		M81186.1	GI:144734			
KEYWORDS		botB gene; neurotoxin type B.				
SOURCE		Clostridium botulinum DNA.				
ORGANISM		Clostridium botulinum				
REFERENCE		1 (bases 1 to 4041)				
AUTHORS		Whelan, S.M., Elmore, M.J., Bodsworth, N.J., Brehm, J.K., Atkinson, T. and Minton, N.P.				
TITLE		Complete nucleotide sequence of the Clostridium botulinum gene encoding the type B neurotoxin				
JOURNAL		Unpublished (1991)				
FEATURES		Location/Qualifiers				
source		1..4041				
gene		/organism="Clostridium botulinum"				
		/db_xref="taxon:1491"				
		57..3932				
		/gene="botB"				

CDS	57..3932	Db	3090	AATAACGCTAAATTTATATATTAATGCTAGACTAGAAATCAAAATACAGATATTAAGGATATA	3149
/gene="botB"		Qy	550	cgtgaagtctatcgtaaacggtgaaatcatctcaaacctggaacggtgacatcgatcacc	609
/function="vertebrate neurotoxin"		Db	3150	AGAGAAGTTATGCTAATGCTGAAATATATTTAAATAGTGGTGTATAGATAGAACAA	3209
/codon_start=1		Qy	610	cagttcatctggatgaataactcttccatcttcaacacccgaactgtctcagtcaccaatc	669
/transl_table=11		Db	3210	CAATTTATTTGGATGAAATATTTTCAGTATTTTAAATACGGAATTAAGTCAATCAATATT	3369
/product="neurotoxin type B"		Qy	670	gaagaacggtacaagatccagctcttactccgaatacctgaaagactctcgggtaataccg	729
/db_xref="GI:144735"		Db	3270	GAAGAAAGATATAAAATCAATCATATAGCAATATTTAAAGATTTTGGGAAATCTCT	3329
/transl_table=11		Qy	730	ctgagtacacaagaataactatctgttcaatctgctggttaacaagaactcttacatcaaa	789
/db_xref="GI:144735"		Db	3330	TAAATGTAATTAAGAAATATATATGTTTAAATGCGGGGAATAAAATTCATATATATAA	3389
/product="neurotoxin type B"		Qy	790	ctgaagaagactctcgggtggtgaaatcctgactccttccaaatcacaacccagaactct	849
/db_xref="GI:144735"		Db	3390	CTAAGAAAGATTCACCTCTAGTGAATTTTAAACACGTAGCAAAATATAATCAAAATCT	3449
/transl_table=11		Qy	850	aaatcacatcaacacgacgacctgtacatcggtgaaagcttcatcctcogtcgcaaatct	909
/db_xref="GI:144735"		Db	3450	AAATATATAATATATAGATTTATATATGGAGAAAAATTTATTAAGAGAAAGTCA	3509
/product="neurotoxin type B"		Qy	910	aactctcagtcacatcaatgacatcgatcgtgaaagactcacaagaactcactcctggactc	969
/db_xref="GI:144735"		Db	3510	AAATCTCAATCATTAATGATGATATAGTAGTAAAGAAAGATTTATATATATCTAGATTT	3569
/transl_table=11		Qy	970	ttaacactgaacaggaatggcgtgtatcacctcacaagcttctcagaagaagaagaa	1029
/db_xref="GI:144735"		Db	3570	TTTAATTTAATCAAGAGTGGAGATATATACCTATAATATTTTAAAGAGAGGAAGAA	3629
/product="neurotoxin type B"		Qy	1030	agactcttcctggtcgcgatctctgatccgacgaactcacaacccatccagatcaaa	1089
/db_xref="GI:144735"		Db	3630	AAATGTTTTTAAAGTCTTAAAGTCTGATGATGATGATGATGATGATGATGATGATGAT	3689
/transl_table=11		Qy	1090	gaatcagcagaacacccgactcctctgacgctcttccagcgtcttccagaagaagaactc	1149
/db_xref="GI:144735"		Db	3690	GAATATGATGAACAGCAACATATAGTGTCTGCTTTTAAAGAGATGAAGAAAGT	3749
/product="neurotoxin type B"		Qy	1150	actgacgaactcgtctgctggtggtatccaccgtttctacgaactctggtatctcgaa	1209
/db_xref="GI:144735"		Db	3750	ACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT	3809
/transl_table=11		Qy	1210	gaatcagaagactcctctgcatctcctcaaatggtacactgaaaggaagttaaacgcaaacg	1269
/db_xref="GI:144735"		Db	3810	GAGTATAAGAGATTTATTTTGTATAAGTAAATGTTTAAAGAGAGGTAAGAAAGCA	3869
/product="neurotoxin type B"		Qy	1270	tacaacctgaactcgtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt	1329
/db_xref="GI:144735"		Db	3870	TATAATTTAAATTTGGATGATTAATGCGCATTTTATTCCTTAAAGATGAAGGGTGAAGT	3929
/transl_table=11		Qy	1330	tagtaagaat	1339
/db_xref="GI:144735"		Db	3930	TAATATACT	3939
RESULT	2				
AF295926					
LOCUS	AF295926				
DEFINITION	Clostridium botulinum neurotoxin type B gene, complete cds.				
ACCESSION	AF295926				
VERSION	AF295926.1				
KEYWORDS	GI:15419707				
SOURCE	Clostridium botulinum.				
ORGANISM	Clostridium botulinum				
REFERENCE	1 (bases 1 to 3876)				
AUTHORS	Kirma,N., Ferreira,J.L. and Baumstark,B.R.				
TITLE	Characterization of six type A strains of Clostridium botulinum				

BASE COUNT	1679 a	383 c	645 g	1334 t
ORIGIN				
Query Match	55.4%;	Score 742.8;	DB 1;	Length 4041;
Best Local Similarity	72.4%;	Pred. NO. 1.1e-170;		
Matches 963;	Conservative	0;	Mismatches 367;	Indels 0;
Gaps 0;				
Qy	10	atggccacaataacatacctcgaacatcctgaaacataatcctcgaacacggtggtacaaa	69	
Db	2610	ATGTTTAAATTAATAAGCAAAATTTAAATAATATATATCTTAAATTAAGATATAAG	2669	
Qy	70	gacaacatctgatcgtctggttacggtgctgctgaaagtgaagatacagcaggtgt	129	
Db	2670	GATAATAATTAATGATTTATCAGSATATGGGCAAGGTAGAGGTATATGATGAGTCT	2729	
Qy	130	gaactgaatgacaagaacagtcacactcctctcgcgtacactcctcgaacggtgt	189	
Db	2730	GAGCTTAATGATAAAATCAATTTAAATTAACCTAGTCTAGCAAAATAGTAAGATGAGTG	2789	
Qy	190	actcagaatcagaacatcctcgaactcgtatctcggactctctggttccctctg	249	
Db	2790	ACTCAAAATCAGAAATATATATTAAGTGTCTGCTTAAATGATTTAGCGTGTGTTGG	2849	
Qy	250	attcgtatcccaataacagaacgacggtatccagaattacatccacaatgaatacacc	309	
Db	2850	ATAAGAATACCTAAATATAAGAAATGATGATACAAATATATATCAATCAATATACA	2909	
Qy	310	atcataactcgaatgaagaataactcctggttggaagatcctcctcgcggtgaacggtatc	369	
Db	2910	ATAATTAATTTGATGAAAAATTAATTCGGGCTGGAATAATCTATTAGGGGTAAAGGATA	2969	
Qy	370	attcgtactcgtatcgaatcgaacggtgaacacacactcgtattctcgaatacacaatc	429	
Db	2970	ATATGAGCTTTAATGATATAAATGGAACCAATCGGTATTTTGAATATATACATA	3029	
Qy	430	cgtgaagacatctcgaatacactcgtggttctcgttaccatcaccataacactcgt	489	
Db	3030	AGAGAAGATATATCAGATATATAAATAGATGTTTTTGTAACTATTTACTTAATTTG	3089	
Qy	490	aacaatgtaaaatcacaacggtgaactggaatcctcgaatacagacatcaagaacatc	549	


```

ACCESSION   AF300465
VERSION     AF300465.1  GI:15982936
KEYWORDS    Clostridium botulinum.
SOURCE      Clostridium botulinum.
ORGANISM    Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
            Clostridium.
REFERENCE   1 (bases 1 to 3876)
AUTHORS    Kirma, N., Ferreira, J.L. and Baumstark, B.R.
TITLE      Characterization of six type A strains of Clostridium botulinum
            that contain type B toxin-gene sequences
JOURNAL    Unpublished
REFERENCE   2 (bases 1 to 3876)
AUTHORS    Kirma, N., Ferreira, J.L. and Baumstark, B.R.
TITLE      Direct Submission
JOURNAL    Submitted (28-AUG-2000) Department of Biology, Georgia State
            University, P.O. Box 4010, Atlanta, GA 30302-4010, USA
FEATURES    Location/Qualifiers
             1..3876
             /organism="Clostridium botulinum"
             /isolate="588"
             /db_xref="taxon:1491"
             /country="USA: Ohio"
             /note="isolated from stool sample in 1976
             type: Ab"
             1..3876
             /codon_start=1
             /transl_table=11
             /product="type B cryptic neurotoxin"
             /protein_id="AAL11498.1"
             /db_xref="GI:15982937"
             /translation="MPVTINNFNDPIDNNIIMPEPPFARGMGRYKAFKIDTRIW
             IIPRYTGVKYPEDFNKSGSIFNRDCEYVDYNTNDKKNIFLOPMKLFNRKIK
             PLGKLELLEINGPILGDRVRPLEEETHIASVTYVKNLISNPFCEVERKKGIFANLII
             FGQPVLENETIEDIGIQNFASREGFGGQMFCPEYVSVFNWVNGKASFNRR
             GYFSDPALILHMLHIVHLGVIKLVNDLPVNPNEKFFQMSQDIAQAEELYTFGGD
             PSIIPTSDKSIYDKVLQNFHGVIVDRNLKVLVCISDPNININIKFQKDKYKEVEDS
             EGYSDIVESFGLKSLMFQFTENIAENYKIKRASYSFSDSLPPVKIKNLIDNEYI
             TIEGFNSDKMEKRYGONKAIKQAEYSKEHLAVYKIQMKSVKAPGICIDVD
             NEDLFTADKNSFSDLSKNERIAYNTQNYIENDFINELIDTLILSLPSEWT
             ESUTDFNYVPYKQPAIKKIFDENTIQYLYSQTFPDLDIRISLTSFDDALLFS
             NKVYSFMDITKANKVVEAGLFAGWQIVNDFVIEANKSSMDIADISLIVPI
             GLALNVGNEAKGNFENAFETAGASILLEPILLIPVGFALLESYIDNKRKLIETI
             NSALTREDKIDMIGYLIQVWLTSTVNTQYTIKEGMYKALNYQAOALEEIKYKNI
             YSEKERNINIDFNVNSKINEGINOININNEINCEYSYLMKMPILAVEKLLDF
             DNTLRKLNLYIDENKLYLIGSAEYKSKVDKYLKTSIPDLSTYNTNTILIEFNKY
             NSDLNLIILNRYRDNKLIIDLSYGAKEVYDGVKLNKQFKLTSANSKRIVTON
             QNLIIFSMFIDFSVFWIRPKYKNDGQIYIHNEYITINCMKNNSGKISIRGNMII
             WTLIDNGIKSVFEYSIKEDISEYINRWFFVTITNNSDNKIYINGKLESHDIED
             IREYIANDEIFKLDGNIDRTQFIWMKYFSIFNTELSQSNIEEYIKQSYSEYLDKF
             GNPLMYKYYMFNAGNKNYSYIKLKDSSVGEILTRSKYNSKNSKYINRDLYIGEKFI
             IRRKNSQSINDIDIVRKEDYIYLDFFNLQEWRYVMYKFKKBEKLEFLAPISDSDEF
             YNTIQIKYEDEPTYSYCOLLEKPKDEESTDEGLIGIHRFYESGIVFKEYKDYFCISKW
             YLKEVKKRPYNSKLGCMWQIPKDEBQWTE"
BASE COUNT  1611 a 367 c 616 g 1282 t
ORIGIN
Query Match      52.7%; Score 706.4; DB 1; Length 3876;
Best Local Similarity 71.0%; Pred. No. 8.8e-162;
Matches 935; Conservative 0; Mismatches 381; Indels 0; Gaps 0;

Qy 16 acaataacaatccgaatccgtgaacataatcatccctgaacctgcgtgaacacacac 75
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2560 AATAAATAATAGGATATTTAAATAATATATATCTTAATTTAAGATATAGGATA 2619

Qy 76 aatctgatcgatctgtctcgttgtaagtgtaagtgataacgacggtgttgaaactg 135
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2620 AAGTTAATAGATTTATCAGGATATGGGCAAGGTAGAGGTATATGATGGGGTCAAGCTT 2679

Qy 136 aatgacaagaaccaggttcaaaactccttccgcgttaacttaagctcgtgttactcag 195
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2680 AATGATAAAATCAATTTAAATTAATTAAGTACAGCAAAATAGTAAGTAGAGTGA 2739

```

```

Qy 136 aatcagaacatcatctctcaactccgtattcctggacttctctctctctctctggtatt 255
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2740 AATCAGATAATCATATATTAATAGTATGTCCTTGATTTAGCGTTAGTTTGGATAAGA 2799

Qy 256 atcccgaaatacaagaacgacggtatccagaattacatcccaatgaataacacatcatc 315
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2800 ATACCTAAATATAAGAATGATGTTATACAAAATTTATTTCAATTAATGAATATAAATAAT 2859

Qy 316 aactgcataagaataaactcgttggaagatctccatcccggttaacacggtatcatctgg 375
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2860 AATTGATGAAAAATAATCTGGATGGAATAATCTATTAGGGTAAATATGATAATATG 2919

Qy 376 actctgatcgatatacaacggttaagaccacaaactctgtatcttcgaatacaacacgtgaa 435
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2920 ACTTTAATGATATAAATGGAATAATCAATCAGTATTTTTTGAATATAGCAATAAGAA 2979

Qy 436 gacatctctgaatacatcaatcgctggtctctcgttaccatcaacaaacacacgaacaa 495
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2980 GATATATCAGAGTATATAAATAGATGTTTGTGTAACATTTACTTAATTAATTCGGATAAC 3039

Qy 496 gctaaaatctacatcaacggttaaaactggaataataacgacatcaaaagacatccgtgaa 555
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3040 GCTAAATTTATTAATGTAAGCTAGAACATACATATAGATATATTAGAGATATAAGAGAA 3099

Qy 556 gtatcgcgttaacggtgaaatcatcttcaactggacggtgacatcgatcgaccagttc 615
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3100 GTTATTGCTAATGATGAATAATATTTAAATAGATGTTAATATAGATAGACACAGTTC 3159

Qy 616 atctggatgaataactctccatcttcaacacccgaaactgtctcagtcacatatacgaa 675
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3160 ATTTGGATGAAATATTTTCACTAGTATTTTAAACGGAAATTAAGTCAATCAAAATTTGAAGAA 3219

Qy 676 cgggtacaagatccagcttactccgaaataacactgaaagactctgggggtaactccgtgatg 735
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3220 ATATATAAAATTCATATATAGCGAAATTTTAAAGAAATTTTGGGAAATCCTTTAATG 3279

Qy 736 tacacaagaataactatactatgttcaatgctggttaacaagaactcttaccataaactgaag 795
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3280 TACAATAAAGAAATATATATGTTTAATCGGGGAATAAAATTCATATATAAATAAAG 3339

Qy 796 aaagactctcgggttgggtgaactcgtgcttcaataatacaacacgaaactcctaataac 855
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3340 AAAGATCTCATCTGAGTGAATTTTAAACGCTAGCAATATAATCAAAATTTCCAATAT 3399

Qy 856 atcaactaccgacactatcatcggtgaaagttcatcccgctcgcaaatcctaactct 915
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3400 ATAAATTTAGAGATTTATATTGGAGAAAATTTTATATAAGAAAGAAAGTCAATTTCT 3459

Qy 916 cagtcacatcaatgatgacatcgtaacgtaaaagaactacatctactcgtgactcttcaac 975
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3460 CAATCTAATATGATGATATAGTTAGAAAAGAAAGATTTATATATCTAGATTTTTTAAT 3519

Qy 976 ctgaatacgggaatggcggtgataacacctacaagtaactcaagaagaagaagaagaagtt 1035
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3520 TTAATCAAGAGTGGAGAGATATATATGATATAATATTTTAAAGAAAGAGAAAGAAATG 3579

Qy 1036 tctcggctccgactctgattccgacgaactctcaacacaccatccagatcaagaataac 1095
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3580 TTTTGTAGTCTCTAATAGTGTCTGATGAGTTTTCATATCTATACAAATAAAGAAATAT 3639

Qy 1096 gacgaacagccgacactcttgcagctgctgttcaagaagaatgaagaatctactgac 1155
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3640 GATGAACAGCCCAACATATAGTTGTGTCAGTTGCTTTTAAAAAAGATGAAGAAAGTACTGAT 3699

Qy 1156 gaaatcggctgatcggttatccacogtttctacgaatctggttatcgattcgaagaatac 1215
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3700 GAGATAGATTGATTGGTATTATCATCGTTCTTACGAATCTCGGAATTTGTTATTAAGAGAT 3759

Qy 1216 aaagactactctgcatctccaaaatgggtacotgaaggaagtttaaacgcgaacccgcacac 1275
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 3760 AAAGATTTATTTGTATAAGTAAATGGTACTTAAAAAGAGGTAAAAAAGGAACCATATAAT 3819

```

Qy	1276	ctgaaactgggttcgaattggcagttccatcccgaagaacgaagtgttgaggacaata	1331
Dd	3820	TCAAAATTGGGATGAATTGGCAGTTTATTCTTAAGATGAAGGTGCAGACTGAATA	3875

RESULT⁴

```

AF300466      3876 bp   DNA       linear    BCT 08-OCT-2001
Clostridium botulinum isolate 593 type B cryptic neurotoxin gene,
complete cds.
ACCESSION     AF300466
VERSION       AF300466.1  GI:15982938
KEYWORDS
SOURCE        Clostridium botulinum.
ORGANISM      Clostridium botulinum
Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae;
Clostridium.
1 (bases 1 to 3876)
Kirma,N., Ferreira,J.L. and Baumstark,B.R.
Characterization of six type A strains of Clostridium botulinum
that contain type B toxin gene sequences
Unpublished
2 (bases 1 to 3876)
Kirma,N., Ferreira,J.L. and Baumstark,B.R.
Direct Submission
Submitted (28-AUG-2000) Department of Biology, Georgia State
University, P.O. Box 4010, Atlanta, GA 30302-4010, USA
Location/Qualifiers
1..3876
/organism="Clostridium botulinum"
/isolate="593"
/db_xref="taxon:1491"
/country="USA: Georgia"
/note="isolated from dog feces in 1976
type: A(B)"
l..3876
/transl_table=1
/product="type B cryptic neurotoxin"
/protein_id="AAL11499_1"
/db_xref="GI:15982939"
/translation="MPVTINFNNDPINDNNIIMPEPPFARGMGRRYYAKFKITDRIW
IPERFTFGYKGFNKKSGIFNRDVCEDYPDLNTNKKNIFLOTMTKLFRNIIRYSK
PLGEKLEMMINGIPFLVDRLRVEEFNTNNTASVTKNLISNPGEVRNKGFIANIL
FGPSDALINMHELHVLHCLGYKIVNDLPVPEKKEFFMOSDTAQOELVTFEGGD
XYSIDSPDKSIYDKVLQNGRVGLRNKVLVCISDPNNINIYNKPKDKKYKFVEDS
EYESIDVSFEKLYIKSLMFPGPDFTIAENYKIKTASFSDLPVKIKLNDLDNII
TIEEGNIKDNNKEYSRQNKAINKQAEEISKHLAVYKIOMKSVKAPGICIDVI
NDLFTIADKNSFSDDLKSNERLAYTNQNNYIENDFSEINELIDLTOLISKEIHPSEN
ESTUDTPVPVYKQPALKIKFTDENTFOYLSOTEPDIRDISLTSSFDALLFES
NKYVSFESMDYIKTANKVPEAGLFAGVKQIVNDPVIANKSSTMDKDIADISLVPIY
GLALNVGTAKAGNFENAEIAGASILTEPIPELLIPVGAFLESIEDNNKNIILETI
NSALTARDEKIDWYGLIYAQMILTSTNYFTYTIKEGMKALNYQAQALBEIKYKYN
YSKERSNINDFNDSYLNKNEGINOAIDNNFINFISECSYLMKMKIPLAVEKLDFI
DNPTLRNLLANYIDENKLYLGISAEEKSKVDKYLKTSIPDLSITYNTNTILIEFNKY
NSDILANIILNLRYSRWKIPRYKNDGIONYIHNEYITIINCKMNSGWKISIRGNMI
ONLIINFMSDELRSVFYSKEDISEYINRWFFVTITNSDNAKIYINGKLESHDIRI
WTUIDINGKISVFPYSKEDISEYINRWFFVTITNSDNAKIYINGKLESHDIRI
IREVIANDELIIFKLDGNIDRTOPIMWKYSIFNETLSQSNIIEYIQSYSEYKDFOF
GNPLYMKTETMYFNAGNKSNYIKLKDDSSYGELLTRSKSQNSKYNTRYRDLYIGEKFI
IRKNSQSINDOPTYKVEDYIYLDQFNLNQEWRVVMKYFKKEEKLFLAPISDSFDE
YNTIOIKYEDEOPTYSCOLLFFKDDSESTDIEGLIGHREYVESIGVFKEYKDYPICISK
YLKEVXRKPYNSKLGCNWQFIPKDEGWE"
```

BASE COUNT 1611 a 367 c 616 g 1282 t

ORIGIN

CDS

JOURNAL

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

SOURCE

Query Match 52.7%; Score 706.4; DB 1; Length 3876;

Best Local Similarity 71.0%; Pred. No. 8.ee-162;

Matches 935; Conservative 0; Mismatches 381; Indels 0; Gaps 0;

Qy 16 acaaatcataattccgaaattcttgaacaaatatcatctgaacctgcattacaaagacaac 75

Qy	316	aactgatgaagaataaactctggttggaagatctccatccggtgtaaacgcgtatcatcgtg	375
Db	336	AAATGCAATAAAATATCTGGCTGGAAATATATCTATTAGGGTAATAGGTAATATG	395
Qy	376	actctgatcatatacaacggtgaagaccacaaactctgtattcttcgaatacaaacatcogtga	435
Db	396	ACTTTAACTGATATAAATGGAAAAACCAATCAGTATTTTTTGAATATAGCATAGAGAA	455
Qy	436	gacatctctgaatacaataaactctggttcttcgtttaccataccacaataaacctgaacat	495
Db	456	GATATATCAGACTATATAAATAGATGCTTTTGTGAATATATACATAAATTCGGATAAC	515
Qy	496	gctaaatctacatacaacggtgaactgaacttaataacgcacatcaaaagacatccgtga	555
Db	516	GCTAAATTTATATTAATGGTAAGCTAGAAACAATATAGATATTAAGATATAGGAGAA	575
Qy	556	gttatcgtaacggtgaatactctcacaactctcgaactctggtgacatcgatccaccgttc	615
Db	576	GTTATTCGTAATAGTGAATAATATTTAAATTAGATGCTGATATAGATAGAACACAAATTT	635
Qy	616	atctgatgaataactctccatctcacaacccgaactgtctcagtcacaataatcgaaaga	675
Db	636	ATTGGATGAATAATTTTCACTATTTTAAACAGAAATTAAGTCAATCAAAATATTAAGAA	695
Qy	676	cgttacaagatccagcttactccgaatacctcgaagactctcgtggttaactccgctgatg	735
Db	696	ATATATAAATTCATCATATATAGCAATATTTAAAGATTTTGGGGAAATCCTTTAATG	755
Qy	736	tacaacaagaataactataatgttcaatctggtgaacagaactcttatacaaacgaag	795
Db	756	TACAATAAGAATAATATATGTTTAAATCGGGGAATAAAATTCATATATTAACATAAG	815
Qy	796	aaagactccggttggaactcgtactcgttcacaatacaaacacgaactcttaataac	855
Db	816	AAAGATTTCATCTGTAGGTGAATTTTACACGCTAGCAAAATTAATCAAAATTCCAATTAT	875
Qy	856	atcaactccgacactgtacatcgttgaaagtgcacatccgtcgcgaactcctaactct	915
Db	876	ATAAATATAGAAATTTATATATCGGAGAAAAATTTATATAAGAGAAAGTCAAAATCT	935
Qy	916	cagtcacaatgatgacatcgtaacgaagaagactacatctacctggactctctcaac	975
Db	936	CAATCTATAAATGATGATATAGTTAGAAAAGAAAGATTATATATATCTAGATTTTTTAAT	995
Qy	976	ctgaatcaggaatgctgtatatacctcacaagtacttcaagaagaagaagaagaagctt	1035
Db	996	TCAATCGAGAGTGGAGAGTATATGCCCTATAAAGATTTTAAAGGAAGAGAAAAAATG	1055
Qy	1036	ttcctggctccgatctctgattccgacgaactctcaacacaccatccgatcaagaatac	1095
Db	1056	GTTTTAGCTAATATATATGATCTTAATCAATTTTACAAACTATACAAATAAAGAAAT	1115
Qy	1096	gacgaacagccgacactctctgacgctgctgttcaagaagaatgaagaatctactgac	1155
Db	1116	GATGAACAGCCCAACATATAGTTGTCATTAATCTTTTAAAAAAGATGAAGAAGTACTGAT	1175
Qy	1156	gaactcgtctgatcgttatccacogttcttaccgaactcgttatcgttatcgaagaatac	1215
Db	1176	GAGATAGATTGATTGGTATTATCATCGTTTTTACGAATCTGGAATTTGATTTAAAGATTAT	1235
Qy	1216	aaagactctctgatcctcacaatggttacctgaaggaagttaaacgcacacgcacac	1275
Db	1236	AAAAATTTATTTTGTATAGTAATAGGTACTTAAAGAGGTAAGAAAGAACCATATATAC	1295
Qy	1276	ctgaactgggttggaattggcagttcatcc	1306
Db	1296	CCCAATTTAGGTGTAATTTGGCAGTTCATTC	1326
RESULT 10			
CBONTB			
LOCUS			

DEFINITION	C.botulinum type B gene for neurotoxin.
ACCESSION	X71343
VERSION	X71343.1 GI:296148
KEYWORDS	bont/B gene; botulinum neurotoxin type B; neurotoxin type B.
SOURCE	Clostridium botulinum.
ORGANISM	Clostridium botulinum Bacteria; Firmicutes; Bacillus/Clostridium group; Clostridiaceae; Clostridium.
REFERENCE	1 (bases 1 to 4051)
AUTHORS	Hutson,R.A., Collins,M.D., East,A.K. and Thompson,D.E.
TITLE	Nucleotide sequence of the gene coding for non-proteolytic Clostridium botulinum type B neurotoxin: comparison with other clostridial neurotoxins
JOURNAL	Curr. Microbiol. 28 (2), 101-110 (1994)
MEDLINE	94122659
REFERENCE	2 (bases 1 to 4051)
AUTHORS	Hutson,R.A.
TITLE	Direct Submission
JOURNAL	Submitted (06-APR-1993) R.A. Hutson, AFRC Institute of Food Research, Reading Laboratory, Microbiology Dept., Earley Gate, Whiteknights Road, Reading, RG6 2EF, UK
FEATURES	Location/Qualifiers
source	1..4051 /organism="Clostridium botulinum" /strain="Exlund 17B ATCC25765" /isolate="type B" /db_xref="taxon:1491" /clone="pCBOTB1-8" 83..91 97..3972 /gene="bont/b" 97..3972 /gene="bont/b" /codon_start=1 /transl_table=11 /product="Bont/B" /protein_id="CAA50482.1" /db_xref="GI:296149" /db_xref="SPTREMBL:Q08077" /translation="MPVTINNFNDPIDNDNIIMPEPPFARGTGRYKAFKITDRIV IIPERYTGGKPEDFNKSGIFNRDVCYEYDPIYNTNDKKNIFQTLIKLFNRKSK PLGKLEMIINGIPYLGDRVPLEEFNTIASVTNKLINSGVEKRRKGIFANLI FGPGVLENENTIDIGQNHFAEGGIMQKPCPEYVSVFNVNKSGASIFNRR GYSDPALILMHELIHVLGLYKQVDDLPVNEKPFQMSQTDIOAEYLFQGD PSITSPTDSKSIYDKVQNFRIIVDRNLKVLVCIISDPNININYNKPKDYKFVEDS EKGYSIDVESNKLYKSLMGFTETINAEYKIKTRASYFSDSLPPVKIKLLDNEIY TIEGFNIDKRMKQYRQNKAKQKAYEISEKHLAVYKIQMCKSVKVGICIDVD NENLFFADKNSFSDLSKNRVEYNTQNNYIGNDFPINELILDLSKILPSENT ESLDFNDVDPVEYKPAIKKVTDENTIFQYLSQTFPLNIRDISLTSFDDALLS SKVYSFSDMYIKTKANKVWEAGLVAGVQVDDPVEIANKSSTMDKIDSLIYPI GLALNVGDETAKGNFESAFETAGSILLEFIPELLIPVGVFLLESYIDNKNILKTI DNALTKRVEKIDMYGLIYAQWLSVTNTQFYTIKGMVKALNYQAQALEELIKYKNI YSEKSNININFINDKLNDGVEDESKVDKYLKTIIPFDLSYSTNIEILIKIFNKY DNTLKLLNLYIDENKLYIGSVEDESKVDKYLKTIIPFDLSYSTNIEILIKIFNKY NSEILNINIILNRYDNLIDLSGYKAVEYDGVKLDKQFKLTSSADSKIRVTQN QNIIFNMFDSVFSFWRIRPKYRNDIOQNTIHNEYTIINCMKNSGKWSIRGRNII WTLIDNGTKTSVFVEYNIREDISEYINRFEFTITNLDNAKIVINCTLNSMDIKD IGEVIYNGEITPKLDGVDVDTQFIWMKYPSIFNTQLNQSNIKEIYKIQSYEYLKDF GNPWMYKNEYTFMAGNSYIKLVKSSVGEILIRSKYNSNINIRNTNLYIGKFI IRRESQSINDDIVRKEDYIHLDLVHHEEMRVYAYKFKQEKEKFLSLISDSNEF YKTIKEIYDEQPSYCOLLEKDEESTDDIGLIGIHRFYESGVLLRKKYKDYFCISKV YLKBVKRKYKSNLGCNWFIPKDEGWTE"
RBS	
gene	
CDS	
BASE COUNT	1679 a 384 c 645 g 1343 t
ORIGIN	


```
Db 2554 ATGTTTAAATATATATACGAAATTTTAAATTAATATATATCTAAATTTAAAGATAAAG 2613
QY 70 gacacaaatcgtatcgtctgtctacggtgctaaagttaaaatcgtatcgtacggtgtt 129
Db 2614 GATAATTAATTAATAGATTTATCAGATATGGGCGAAAGTAGAGGTATATGATGAGTC 2673
QY 130 gaactgaatgacagaacacaggttcaactgacacctctccgctcaactcgaactccgtgtt 189
Db 2674 GAGCTTAATGATATAAATCAATTTAAATTAACCTAGTTCAGCAAAATAGATAGAGTG 2733
QY 190 actcagaatcagaacatcattcctcaactcogtattcctggactctctgttctctctgg 249
Db 2734 ACTCAAAATCAGAAATATCATATTAATTAATAGTGTCTCTCTGATTTTACGCTTACCTTTGG 2793
QY 250 attcgtatcccgaaatcacagaacgagctgtatccagaatcattcacatcgaactcacc 309
Db 2794 ATAAGAAATACCTAAATATAGAATGATGGTATACAAATTTATATTCATAATGAATATACA 2853
QY 310 atcatcaactgcatgagaataaactctgttggaagatctccatccggtgaaccgtatc 369
Db 2854 ATAATTAATTTGATGAAATAATTTCCGGCTGGCAAAATATCTATTAGGGGTAAATAGATA 2913
QY 370 atctggactctgcatgatcatcaacggttaagaccacaaatctgttatcttcggaatacaacatc 429
Db 2914 ATATGGACTTTAATTTGATATAAATTTGGAATAACCAAAATCGGTATTTTGTGAATAATAACA 2973
QY 430 cgtgagacatctcgtgaatcacatcgtggttctgttctgttaccatcaccataaactg 489
Db 2974 AGAGAAGATATATCAGAGATATATAAATAGATGTTTTTGTGAATTAATTAATAATTTG 3033
QY 250 attcgtatcccgaaatcacagaacgagctgtatccagaatcattcacatcgaactcacc 309
Db 2794 ATAAGAAATACCTAAATATAGAATGATGGTATACAAATTTATATTCATAATGAATATACA 2853
QY 310 atcatcaactgcatgagaataaactctgttggaagatctccatccggtgaaccgtatc 369
Db 2854 ATAATTAATTTGATGAAATAATTTCCGGCTGGCAAAATATCTATTAGGGGTAAATAGATA 2913
QY 370 atctggactctgcatgatcatcaacggttaagaccacaaatctgttatcttcggaatacaacatc 429
Db 2914 ATATGGACTTTAATTTGATATAAATTTGGAATAACCAAAATCGGTATTTTGTGAATAATAACA 2973
QY 430 cgtgagacatctcgtgaatcacatcgtggttctgttctgttaccatcaccataaactg 489
Db 2974 AGAGAAGATATATCAGAGATATATAAATAGATGTTTTTGTGAATTAATTAATAATTTG 3033
QY 490 acaatgctcaaatctcacatcaacggttaactggaactcctaataccgacatcaagaacatc 549
Db 3034 AATAACGCTAAATTTATATTAATGATGTAAGCTAGATCAATCAATACAGATATAAAGATATA 3093
QY 550 cgtggaagtctacgacggtgaataatctctcacaactggacggtgacatcgatcgtaacc 609
Db 3094 AGAGAAGTTTATTCCTAATGCTGAATAATATTTAAATTTAGATGGTGTATATAGATAGAACA 3153
QY 610 cagttcatctcgtggaataactctcctcaactcctcaacacggaactgctcagtcaccaatc 669
Db 3154 CAATTTATTTGGATGAATATATTTTCACTATTTTAAACGGAATTAAGTCAATCAATATATT 3213
QY 670 gaagaacggtacagaatccagctctactccgaataactgaaagactctcgggtaactccg 729
Db 3214 GAAGAAAGATATAAATTTCAATCATATAGCGAATATTTAAAGATTTTGGGGAATTCCT 3273
QY 730 ctgagtacaaagaataactatgttcaatgctggtaacagaactcttcaatcaaaa 789
Db 3274 TTAATGTACAATAAAGAATATATATGTTTAATGCGGGGAATAAAAATTCATATATATAA 3333
QY 790 ctgaagaagaactctcgggtgtgtaactcctgactcgttccaaatcacacagaactct 849
Db 3334 CTAAAGAAGATTCACCTGTAGTGAATTTTAAACAGTAGCAAAATATATCAATCAAAATTC 3393
QY 850 aaatcatcaactaccgacacctgtacatcgttggaagttcatcctcogtccgaactct 909
Db 3394 AAATATATAAATTAAGAGATTTATATTTGGAGAAAAATTTTATTAAGAAGAAAGTCA 3453
QY 910 aactctcagtcacatcgaatgatgacatcgtaacgtgtaagaagaactcacatctacctgga 965
Db 3454 AATTCCTCAATCTATAAATGATGATATAGTAGAAAAAGAAGATTTATATATATCTAGA 3509
```

RESULT 12

```
BD009886
LOCUS BD009886 3509 bp DNA linear PAR 31-JAN-2002
DEFINITION Recombinant toxin fragments.
ACCESSION BD009886
VERSION BD009886.1 GI:18638259
KEYWORDS JP 2001502890-A/10.
SOURCE unidentified.
ORGANISM unidentified
```

```
unclassified.
1 (bases 1 to 3509)
Shone,C., Quinn,C.P. and Foster,K.A.
Recombinant toxin fragments
Patent: JP 2001502890-A 10 06-MAR-2001;
MICROBIOLOGICAL RESEARCH AUTHORITY CAMR,THE SPEYWOOD LABORATORY LTD
OS Unidentified
PN JP 2001502890-A/10
PD 06-MAR-2001
PF 22-AUG-1997 JP 1998510524
PR 23-AUG-1996 GB 9617671.4,13-DEC-1996 GB 9625996.5 PI
CLIFFORD CHARLES SHONE,CONRAD PADRAIG QUINN,KEITH ALAN FOSTER PC
CI2N15/31,CI2N1/21,CI2P21/02,C07K14/33,A61K38/16,A61K39/08 CC
Strandedness: Double;
CC Topology: Linear;
FH Key Location/Qualifiers
FT CDS 1..3509.
source I..3509
/organism="unidentified"
/db_xref="taxon:32644"
BASE COUNT 1468 a 340 c 548 g 1153 t
ORIGIN
```

```
Query Match 39.0%; Score 522.4; DB 6; Length 3509;
Best Local Similarity 71.7%; Pred. No. 7,2e-117;
Matches 685; Conservative 0; Mismatches 271; Indels 0; Gaps 0;
```

```
QY 10 atggccacaacaaatcacattccgaataactcgaacataatcatcctgaacctcgttacaacaa 69
Db 2554 ATGTTTAAATATATATACGAAATTTTAAATTAATATATCTAAATTTAAAGATAAAG 2613
QY 70 gacacaaatcgtatcgtctgtctacggtgctaaagttaaaatcgtatcgtacggtgtt 129
Db 2614 GATAATTAATTAATAGATTTATCAGGATATGGGCGAAAGTAGAGGTATATGATGAGTC 2673
QY 130 gaactgaatgacagaacacggttcaactgacacctctccgctcaactcgaactcgtgtt 189
Db 2674 GAGCTTAATGATATAAATCAATTTAAATTAACCTAGTTCAGCAAAATAGATAGAGTG 2733
QY 190 actcagaatcagaacatcattcctcaactcogtattcctggactctctgttctctctctgg 249
Db 2734 ACTCAAAATCAGAAATATCATATTAATTAATAGTGTCTCTCTGATTTTACGCTTACCTTTGG 2793
QY 250 attcgtatcccgaaatcacagaacgagctgtatccagaatcattcacatcgaactcacc 309
Db 2794 ATAAGAAATACCTAAATATAGAATGATGGTATACAAATTTATATTCATAATGAATATACA 2853
QY 310 atcatcaactgcatgagaataaactctgttggaagatctccatccggtgaaccgtatc 369
Db 2854 ATAATTAATTTGATGAAATAATTTCCGGCTGGCAAAATATCTATTAGGGGTAAATAGATA 2913
QY 370 atctggactctgcatgatcatcaacggttaagaccacaaatctgttatcttcggaatacaacatc 429
Db 2914 ATATGGACTTTAATTTGATATAAATTTGGAATAACCAAAATCGGTATTTTGTGAATAATAACA 2973
QY 430 cgtgagacatctcgtgaatcacatcgtggttctgttctgttaccatcaccataaactg 489
Db 2974 AGAGAAGATATATCAGAGATATATAAATAGATGTTTTTGTGAATTAATTAATAATTTG 3033
QY 490 acaatgctcaaatctcacatcaacggttaactggaactcctaataccgacatcaagaacatc 549
Db 3034 AATAACGCTAAATTTATATTAATGATGTAAGCTAGATCAATCAATACAGATATAAAGATATA 3093
QY 550 cgtggaagtctacgacggtgaataatctctcacaactggacggtgacatcgatcgtaacc 609
Db 3094 AGAGAAGTTTATTCCTAATGCTGAATAATATTTAAATTTAGATGGTGTATATAGATAGAACA 3153
QY 610 cagttcatctcgtggaataactctcctcaactcctcaacacggaactgctcagtcaccaatc 669
Db 3154 CAATTTATTTGGATGAATATATTTTCACTATTTTAAACGGAATTAAGTCAATCAATATATT 3213
```


RESULT 14	AR000029	AR000029	1330 bp	DNA	linear	PAT 04-DEC-1998
LOCUS	AR000029	Sequence 22 from patent US 5736139.				
DEFINITION	AR000029	Sequence 22 from patent US 5736139.				
ACCESSION	AR000029.1	GI:3962560				
VERSION	AR000029.1	GI:3962560				
KEYWORDS	Unknown.					
SOURCE	Unknown.					
ORGANISM	Unknown.					
REFERENCE	1. (bases 1 to 1330)					
AUTHORS	Kink,J.A., Thailay,B.S., Stafford,D.C., Firca,J.R. and Padhye,N.V.					
TITLE	Treatment of Clostridium difficile induced disease					
JOURNAL	Patent: US 5736139-A 22 07-APR-1998;					
ATUES	Location/Qualifiers					
source	1..1330					
BASE COUNT	400 a	339 c	246 g	345 t		
ORIGIN						
Query Match	19.68;	Score 263;	DB 6;	Length 1330;		
Best Local Similarity	53.9%;	Pred. No. 1.5e-53;				
Matches 639;	Conservative	0;	Mismatches 525;	Indels 21;	Gaps 4;	
Qy	15	caacaatacaattccgaatacctgaacaatatcatctgaacctgcgtttacaagacaa	74			
Db	24	CAC TGAATACATCAAGAACATCATCAATACCTCATCTCGAACCCTGCGCTACGAATCCAA	83			
Qy	75	caatctgatcgatctgtctgtttacggtgcgtcaaaagttgaagtatacgcagcggttt-----	129			
Db	84	TCACCTGATGACCTGTCTCGTACGGCTTCCAAATCAACATCGGTTCTAAGTTAACTT	143			
Qy	130	-gaactgaatgacaagaacacagtttcaaaactgacctcttcgcgttaacttaagatacgcgtgt	188			
Db	144	CGATCCGATGCACAGAATCATGATCCAGCTGTTCAATCTGGAATCTTCCAAAATCGAAAGT	203			
Qy	189	tactcagaatcagaacatcatcttcaactccgattcctgcgttaacttctgttcttctctgt	248			
Db	204	TATCCTGAAGAAATGCTATCGTATACAACTATATGATGAGAAACATTTCCACCTCCTCTCTG	263			
Qy	249	gattcgtatccggaatatacaagaacgacggttatccagaattacatccacaatgaatacac	308			
Db	264	GATCGGTATCCGAAATACTTCAACTCC-----ATCTCTCTGACAAATGAATACAC	314			
Qy	309	catcatcaactgcgatgaagaaataactctggttggaagatctccatccgcgcgttaaccggtat	368			
Db	315	CATCATCAACTGCATGGAACAAATTTCTGTTGGAAGATATCTCTGAACTACGGTGAAT	374			
Qy	369	catctgacctctgatcatatcaacggttaacacaaatctgtattcttcggaatacaacat	428			
Db	375	CATCTGGACTCTCAGGACATCTCAGGAATCAACACGGGCTGTTGTTTCAATACTCTCA	434			
Qy	429	ccgtgaagacatctctgaatacatcaatcgcgtggtcttcgtttaccatccaccaataa---	485			
Db	435	GATGATCAACATCTCTGACTACATCAATCGTGGATCTCGTTACCATCACCACAATCG	494			
Qy	486	cctgaacaatgcttaaaatcttatacaacggttaaaactggaatctataatccgcacataaaga	545			
Db	495	TCTGAATAACTCCAAAATCTCATCAACGGCGCTGATCGACCAAGAACCGATCTCCAA	554			
Qy	546	catccggtgaagattatcgctaaacggtgaaatcatcttcaaacctgaacgtgcacatgcctc	605			
Db	555	TCTGGGTAAACATCCACCGCTCTTAATAACATCATGTTTCAAACTGGACGGTGTCTGTGACAC	614			
Qy	606	taccaggttcatctggatgaaataacttctccatcttcaacacgaactgtctcagtcacaa	665			
Db	615	TCACCGCTACATCTGGATCAATATCTCAATCTGTTTCGACAAAACATGACACGAAAGA	674			
Qy	666	tatcgaagaacggttacaagatccaggtcttactccgaatcacctgaaagacttctgtgggttaa	725			

Db 204 TATCCTGAAGAAATGCTATCGTATATAAATCTATGTACGAAAACTTCTCCACCTCCTTCTG 263
Qy 249 gattcgtatcccaagaatacagaacgagcggtatccagaattacataccacaataacacac 308
Db 264 GATCCGTATCCGGAATACATCTCAACTCC-----ATCTCTCTGAACAATGAATACAC 314
Qy 309 catcatcaactgatgaagaataactctggttggaagatctccatccgcggttaaccggtat 368
Db 315 CATCATCAACTGCATGGAACAATCTGTTGGTGAAGATATCTCTGAACACGCGTGAAT 374
Qy 369 catctgactctgatcgatatacaacggttaagaccacaatactctgtattcttcgaatacaacat 428
Db 375 CATCTGGACTCTCGAGGACTCTCAGGAATCAACACGCGTGTGTTATTCAAATACTCTCA 434
Qy 429 ccgtgaagaatactctgaatacatcatcgaatcgctggttcttccatccccaataa--- 485
Db 435 GATGATCAACATCTCTGACTACATCAATCGTGGATCTCGTTACCATCCACAATCG 494
Qy 486 cctgaacaatgctaaatactacatacaacggttaaaactggaatcctaataccgacatacaaga 545
Db 495 TCTGAATACTCCAAATCTACATCAACGCGCTCTGATCGACCGAAGACGATCTCCAA 554
Qy 546 catccgtgaagtattatcgtaacggtgaatactcttcaaaactggacggtgacatcgatcg 605
Db 555 TCTGGTAACATCCACGCTTCTAATAACATCATGTTCAAACTGGACGTTGTCGTGACAC 614
Qy 606 taccagttcatctggtgaataactctccatctccatcttcaacacggaactgtctcagtcacaa 665
Db 615 TCACCGTACATCTGGATCAAAATCTCAATCTGTTGACAAAGAAAGTGAACGAAAAAGA 674
Qy 666 tatcgaagaacggtacaaactccagttctactccgaataactcgaagaactctcggggttaa 725
Db 675 AATCAAGAGCTGTACACACACAGTCCAAATCTGGTATCTCTGAAGACTCTCGGGGTGA 734
Qy 726 tccgctgatgtacacaaagaataactatgttcaatgtcgtggttaacagaactcttaccat 785
Db 735 CTACCTGCAGTACGACAAACCGTACTACATCTGTAATCTGTACGATCCGAACAATACGT 794
Qy 786 caaactgaagaagaactctccggttggtgaataactcgaactcgtctccaaatacaaccagaa 845
Db 795 TGACGTCAACAATGTAGGTATCCGCGTTTACATGTACCTGAAAGTCCGCGTGGTCTGT 854
Qy 846 ctctaaatacatcaactaccgacactgtacatcgtgaaagttcatccatccgctcgcaa 905
Db 855 TATGACTACCAACATCTACCTGAATCTTCCCTGTACCGTGGTACCAATTCATCATCAA 914
Qy 906 atctaactctcagttccatcaatgatgacatcgtacgtacgttaagaagaactacatctacctgga 965
Db 915 GAAATACGCGTCTGGTAACAAGGACAATATCGTTGCGCAACAATGATCGTGTATATACATCAA 974
Qy 966 ctcttcaactgaatacaggaatggcggtgtatacacactcaagaactcctcaagaagaaga 1025
Db 975 TGTGTAGTTAAGAACAAAGAAATACCGTCTGGTACCAATGCTTCTCAGGCTGGTGTAGA 1034
Qy 1026 agaaaagctttctcgtcggtatcgtgattccgacgaactctacacaccatccaat 1085
Db 1035 AAAGATCTTGTCTCTGGAATCCCGGAGGTTGGTAAATCTGTCTCAGGTAGTTGTAAT 1094
Qy 1086 caaagaatacgaacagacgacactactcttgcagctgctgttcaagaagaagaatgaaga 1145
Db 1095 GAAATCCAAGAACGACCAAGGTA---TCACTAACAAATGCAAAATGAATCTGCAGGACAA 1151
Qy 1146 atctaactgaagaatacgtctgatcggtatccaccggtttctacga 1190
Db 1152 CAATGGTAACGATATCGGTTTCATCGGTTTCCACCAAGTTCAACAA 1196